ZENECA Inc.
Docket No. 70086

#3

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION OF: WOOD ET AL.

SERIAL NO.: 08/669,656 GROUP ART UNIT: UNKNOWN

FILED: JUNE 24, 1996 EXAMINER: UNKNOWN

FOR: ION CHANNEL

Certificate of Mailing (37 CFR 1.8(a)).
I hereby certify that the attached papers or fee is being deposited with the United States Postal Service on the date shown below with sufficient postage as first class mail in an envelope addressed to: Box Missing Part, Assistant Commissioner For Patents, Washington, D.C. 20231.

Liza D. Hohenschutz (Printed Name) September 20, 1996 (Date)

Liza D. Hohenschutz (Signature)

Box Missing Part
Assistant Commissioner For Patents
Washington, D.C. 20231

Sir:

TRANSMITTAL OF PRIORITY DOCUMENT

Submitted herewith is a certified copy of applicants' priority application Serial No. 9513180.1 filed in the United Kingdom on June 28, 1995. Applicants reiterate the claim of priority under 35 USC 119/365 and a prompt acknowledgment of receipt of the priority document is requested.

Respectfully submitted,

ZENECA Inc.

Date: September 20, 1996

By: Liza D. Hohenschutz
Liza D. Hohenschutz
Attorney for Applicant(s)
Registration No. 33,712
Telephone: 302/886-7466

Docket No.: 70086

y:\ldh\70086\priordoc.doc

7008C
SN 08/669,656

The Patent Office

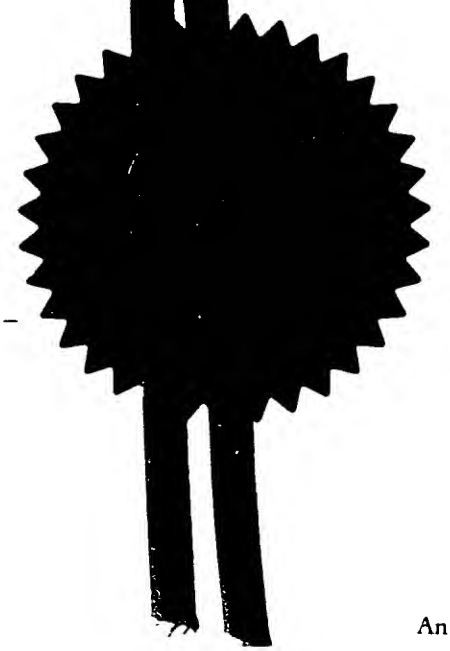
The Patent Office
Cardiff Road
Newport
Gwent
NP9 1RH

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation and Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein together with the Statement of inventorship and of right to grant of a Patent (Form 7/77), which was subsequently filed.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed M.S. Dawe

Dated 1ST JULY 1996

For official use

PATENT OFFICE
28 JUN 1995
RECEIVED BY FAX

9513180.1

29JUN95 E131019-1 002934
P01/7700 25.00

Your reference
PHM.95/41

28 JUN 1995

Notes
Please type, or write in dark ink using CAPITAL letters. A prescribed fee is payable for a request for grant of a patent. For details, please contact the Patent Office (telephone 071–438 4700).

Rule 16 of the Patents Rules 1990 is the main rule governing the completion and filing of this form.

The Patent Office — Request for grant of a Patent

Form 1/77

Patents Act 1977

❶ Title of invention

1 Please give the title of the invention: ION CHANNEL

❷ Do not give trading styles, for example, 'Trading as XYZ company', nationality or former names, for example, 'formerly (known as) ABC Ltd' as these are not required.

❷ Applicant's details

☐ **First or only applicant**

2a If you are applying as a corporate body please give:

Corporate name: UNIVERSITY COLLEGE LONDON

Country *(and State of incorporation, if appropriate)*: ENGLAND

2b If you are applying as an individual or one of a partnership please give in full:

Surname

Forenames

2c **In all cases**, please give the following details:

Address: GOWER STREET
LONDON

UK postcode *(if applicable)*: WC1E 6BT

Country: ENGLAND

ADP number *(if known)*: 7986520002

Warning
After an application for a Patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977 and will inform the applicant if such prohibition or restriction is necessary. Applicants resident in the United Kingdom are also reminded that under Section 23, applications may not be filed abroad without written permission unless an application has been filed not less than 6 weeks previously in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction revoked.

2d, 2e and 2f: If there are further applicants please provide details on a separate sheet of paper.

☐ **Second applicant** *(if any)*

2d If you are applying as a corporate body please give:

Corporate name

Country *(and State of incorporation, if appropriate)*

2e If you are applying as an individual or one of a partnership please give in full:

Surname

Forenames

2f **In all cases**, please give the following details:

Address

UK postcode *(if applicable)*

Country

ADP number *(if known)*

⊛ *An address for service in the United Kingdom must be supplied*

Please mark correct box

⊛ Address for service details

3a Have you appointed an agent to deal with your application?

Yes [X] No [ ] ➡ *go to 3b*

↓ *please give details below*

Agent's name John Richard MACK

Agent's address ZENECA Limited
15 Stanhope Gate
LONDON

Postcode W1Y 6LN

Agent's ADP number ~~01192087002~~ 6799738001

3b: If you have appointed an agent, all correspondence concerning your application will be sent to the agent's United Kingdom address.

3b If you have not appointed an agent please give a name and address in the United Kingdom to which all correspondence will be sent:

Name

Address

Postcode

Daytime telephone number *(if available)*

ADP number *(if known)*

❶ Reference number

4 Agent's or applicant's reference number *(if applicable)* PEM.95/41

❷ Claiming an earlier application date

5 Are you claiming that this application be treated as having been filed on the date of filing of an earlier application?

Please mark correct box

Yes ☐ No ☒ ➡ *go to 6*

↓ *please give details below*

❑ number of earlier application or patent number

❑ filing date

(day month year)

❑ and the Section of the Patents Act 1977 under which you are claiming:

Please mark correct box

15(4) (Divisional) ☐ 8(3) ☐ 12(6) ☐ 37(4) ☐

❸ Declaration of priority

❸ If you are declaring priority from a PCT Application please enter 'PCT' as the country and enter the country code (for example, GB) as part of the application number.

Please give the date in all number format, for example, 31/05/90 for 31 May 1990.

6 If you are declaring priority from previous application(s), please give:

Country of filing	Priority application number *(if known)*	Filing date *(day, month, year)*

The answer must be 'No' if:
- *any applicant is not an inventor*
- *there is an inventor who is not an applicant, or*
- *any applicant is a corporate body.*

6 Inventorship

7 Are you (the applicant or applicants) the sole inventor or the joint inventors?

Please mark correct box

Yes [ ] No [X] ➡ ***A Statement of Inventorship on Patents Form 7/77 will need to be filed (see Rule 16).***

Please supply duplicates of claim(s), abstract, description and drawing(s).

8 Checklist

8a Please fill in the number of sheets for each of the following types of document contained in this application.

Continuation sheets for this Patents Form 1/77			
Claim(s)		Description	32
Abstract		Drawing(s)	33

8b Which of the following documents also accompanies the application?

Please mark correct box(es)

Priority documents *(please state how many)*	
Translation(s) of Priority documents *(please state how many)*	
Patents Form 7/77 – Statement of Inventorship and Right to Grant *(please state how many)*	
Patents Form 9/77 – Preliminary Examination/Search	
Patents Form 10/77 – Request for Substantive Examination	

You or your appointed agent (see Rule 90 of the Patents Rules 1990) must sign this request.

9 Request

I/We request the grant of a patent on the basis of this application.

Please sign here ➡

Signed [signature] Date 28 / 6 / 95 (day / month / year)

A completed fee sheet should preferably accompany the fee.

Please return the completed form, attachments and duplicates where requested, together with the prescribed fee to either:

❑ ***The Comptroller***
The Patent Office
Cardiff Road
Newport
Gwent
NP9 1RH

or

❑ ***The Comptroller***
The Patent Office
25 Southampton Buildings
London
WC2A 1AY

For official use

THE PATENT OFFICE
16 OCT 1995
RECEIVED BY POST

Your reference

Notes
Please type, or write in dark ink using CAPITAL letters.

No fee is required with this form.

Rule 15 of the Patents Rules 1990 is the main rule governing the completion and filing of this form.

If you do not have enough space please use a separate sheet of paper.

When an application does not declare priority at all, or declares priority from an earlier UK application, sufficient additional copies of this form must be supplied to enable the Comptroller to send one to each inventor who is not an applicant.

A

The Patent Office

Statement of inventorship and of right to grant of a Patent

Form 7/77

Patents Act 1977

❶ Application details

1a Please give the patent application number *(if known)*:

9513180.1

1b Please give the full name(s) of the applicant(s):

University College London

❷ Title of invention

2 Please give the title of the invention:

ion channel

❸ Derivation of right

3 Please state how the applicant(s) derive(s) the right to be granted a patent:

EMPLOYER

❹ Declaration

4 I believe the person(s) named overleaf (and on any supplementary copies of this form) to be the inventor(s) of the invention for which the patent application has been made. I consent to the disclosure of the details contained in this form to each inventor named.

General Manager UCLi

Please sign here ➡

Signed

Date 24 (day) 4 (month) 1995 (year)

Please turn over ⇨

084 55 03

Please put the full name(s) and address(es) of the inventors in the boxes below:

Please underline the surnames or family names.

Dr Armen N. ATTOPIAN
Nalbandian 13-7,
Yerevan 375010 ARMENIA 068718260001.
or
Krokuvos 1-8
Vilnius-5
2005 Lithuania

ADP number (*if known*):

Dr. John N. WOOD
50A Chalcot Road,
London NW1 8LS
U.K.
068718340001.

ADP number (*if known*):

ADP number (*if known*):

Please give the names of any further inventors on the back of another form 7/77 and attach it to this form.

Reminder

Have you signed the declaration overleaf?

ION CHANNEL

APPLICANT: University College London
Gover Street
LONDON
WC1E 6BT
England

INVENTORS John Nicholas Wood
50A Chalcot Road
LONDON
NW1 8LS
England

Nationality: British

Armen Norakovitch Akopian
Department of Anantomy and Developmental Biology
University College London
Gower Street
LONDON
WC1E 6BT
England

Nationality: Armenian

ION CHANNEL

The present invention relates to novel voltage-gated sodium channel proteins specific to sensory neurones, to nucleotide sequences capable of encoding these sodium channel proteins, to vectors comprising the nucleotide sequence, to host cells containing these vectors, to cells transformed with the DNA, to screens using the sodium channel proteins and/or transformed cells, to modulators of the sodium channel protein identified using the screen, to complementary stands of the DNA sequence which is capable of encoding the sodium channel proteins and to antibodies specific for the sodium channel proteins.

Voltage-gated sodium channels are transmembrane proteins which cause sodium permeability to increase. Depolarisation of the plasma membrane causes sodium channels to open allowing sodium ions to enter along the electrochemical gradient creating an action potential.

Voltage-gated sodium channels are expressed by all electrically excitable cells, where they play an essential role in action potential propagation. They comprise a major subunit of about 2000 amino acids which is divided into four domains (D1-D4), each of which contains 6 membrane-spanning regions (S1-S6). The alpha-subunit is usually associated with 2 smaller subunits (beta-1 and beta-2) that influence the gating kinetics of the channel. These channels show remarkable ion selectivity, with little permeability to other monovalent or divalent cations. Patch-clamp studies have shown that depolarisation leads to activation with a typical conductance of about 20pS, reflecting ion movement at the rate of 107 ions/second/channel. The channel inactivates within milliseconds (Caterall 1992, Omri et al. 1992, Hille 1994). Sodium channels have been pharmacologically characterised using toxins which bind to distinct sites on sodium channels. The heterocyclic guanidine-based channel blockers tetrodotoxin (TTX) and saxitoxin (STX) bind to a site in the S5-S6 loop, whilst mμ-conotoxin binds to an adjacent overlapping region. A number of toxins from sea anemones or scorpions binding at other sites alter the voltage-dependence of activation or inactivation. Our understanding of the mechanism of action and role of

voltage gated sodium channels has been enhanced by the molecular cloning of several such channels and associated subunits. The electroplax sodium channel was cloned after purification and protein sequencing. The resulting cDNA probes were subsequently used to identify three neuronal sodium channels (types I, II, and III) from rat brain by low stringency hybridisation. The expression of the three genes is developmentally regulated and the type II gene is alternatively spliced to give rise to type II channels in developing brain and type IIa channels that are expressed at higher levels in the adult brain. Three types of sodium chanels have been distinguished by physiological criteria in adult rat dorsal root ganglion neurons (Cafrey et al. 1992, Nowycky 1993, Jeftjina 1994), but in general little is known about the neuron-specific distribution of distinct forms of sodium channels.

We have now found a voltage-gated sodium channel (hereinafter referred to as a sodium channel specifically located in sensory neurones) that is present in sensory neurons but not present in glia, muscle, or the neurones of the sympathetic, parasympathetic, enteric or central nervous systems.

The sodium channel specifically located in sensory neurones shows relative insensitivity to tetrodotoxin (IC50 > 1 micromolar).

Voltage-gated sodium channels that are blocked by nanomolar concentrations of tetrodotoxin are known as tetrodotoxin sensitive sodium channels (Hille 1994) whilst sodium channels that are blocked by concentrations greater than 1 micromolar are known as tetrodotoxin-insensitive (TTXi) sodium channels (Pearce and Duchen 1994).

Preferably the sensory neurones are the neurones of the dorsal root ganglia (DRG) or cranial ganglia.

Most preferably the sensory neurones are the neurones of the dorsal root ganglia.

Preferably the sodium channel is specifically located in rat sensory neurones.

Most preferably the sodium channel has the amino acid sequence set out in **Figure 1a** or a conservative analogue or a splice variant thereof.

A conservative analogue is a protein sequence which retains the biological properties of the sodium channel but differs in sequences by one or more conservative amino acid substitutions. For the purposes of this document a conservative amino acid substitution is a substitution whose probability of occuring in nature is greater than ten times the probability of that substitution occuring by chance (as defined by the computational methods described byDayhoff et al, Atlas of Proteins Sequence and Structure, 1971, page 95-96 and figure 9-10).

A splice variant is a protein product of the same gene, generated by alternative splicing of mRNA, that contains additions or deletions within the coding region (Lewin 1995). Two known splice variants of the original sodium channel are shown in **Figure 1b** and **1c**.

In another aspect the present invention provides a sodium channel specifically located in rat sensory neurones as encosed by the insert deposited in NCIMB deposit number 40744, which was deposited at The National Collections of Industrial and Marine Bacteria, 23 St Machar Drive, Aberdeen AB2 1RY, Scotland, United Kingdom on 27 June 1995 in accordance with the Budapest Treaty.

In another aspect the present invention provides a sodium channel specifically located in human sensory neurones.

The invention also provides a nucleotide sequence encoding a sodium channel or a complementary strand thereof.

Nucleotide sequences may be protein nucleic acid (PNA), DNA or RNA and DNA may be genomic or cDNA.

Preferably, the nucleotide sequence encodes a sodium channel specifically located in rat sensory neurones which is as set out in **Figure 1a** or a complementary strand thereof.

The invention also provides an expression vector comprising a nucleotide sequence as hereinabove defined. In order to effect transformation, DNA sequences containing the desired coding sequence and control sequences in operable linkage (so that hosts transformed with these sequences are capable of producing the encoded proteins) may be included on a vector, however, the relevant DNA may then also be integrated into the host chromosome.

The invention also provides a host cell comprising an

expression vector as hereinabove defined.

Suitable host cells include xenopus laevis oocytes, mammalian cells such as COS-7, HEK 293 cells and NIH303 cells, insect cells especially SF9 or 21 clonal cells derived from Spodoptera fugiperata (in conjunction with baculovirus vectors).

The invention also provides a cell transformed with a nucleotide capable of encoding a sodium channel as hereinabove defined.

The invention also provides a screeing assay for modulators of the sodium channel which is specifically located in sensory neurones wherein the assay comprises contacting a potential modulator with a transformed cell and detecting any change in activity of the sodium channel.

The present invention also provides a modulator which has activity in the screening assay hereinabove defined. Modulators of the sodium channel as hereinabove defined are useful in modulating the sensation of pain. Blockers of the sodium channel will prevent the trasmission of impulses along sensory neurones and thereby be useful in the treatment of acute, chronic or neuropathic pain.

The invention therefore also relates to the use of blockers of the sodium channal as hereinabove defined in the treatment of acute, chronic or neuropathic pain.

Typically a sodium channel consists of a single principal alpha-subunit containing the voltage sensor and sodium permeable pore, and one or two associated subunits (beta-1 and beta-2) which modify the gating characteristics of the channels but are not critical for activity (Catterall 1992).

The present invention also provides antibodies specific for the sodium channels hereinabove defined. The term antibody as used herein includes all Immunoglobulins and fragments thereof which contain recognition sites for antigenic determinants of proteins of the present invention. The antibodies of the present invention may be polyclonal or preferably monoclonal, may be intact antibody molecules or fragments containing the active binding region of the antibody, e.g. Fab or $F(ab)_2$ and can be produced using techniques well established in the art [see e.g. R.A DeWeger et al; Immunological

Rev., 62 p29-45 (1982)].

The antibodies of the present invention may also be of interest in purifying a polypeptide of the present invention and accordingly we further provide a method of purifying a polypeptide of the present invention as hereinabove defined or any portion thereof or a metabolite or degration product thereof which method comprises the use of an antibody of the present invention.

The purification method of the present invention may be effected by any convenient technique known in the art for example by providing the antibody on a support and contacting the antibody with a solution containing the polypeptide whereby the antibody binds to the polypeptide of the present invention. The polypeptide may be released from binding with the antibody by known methods for example by changing the ionic strength of the solution in contact with the complex of the polypeptide/antibody.

Complementary strands of the nucleotide sequences as hereinabove defined can be used in gene therapy [US 5,399,346].

For example, the cDNA sequence or fragments thereof could be used in gene therapy strategies to down regulate the receptor. Antisense oligonucleotides or an antisense contruct driven by a strong constitutive promoter expressed in the target sensory neurones would be delivered either peripherally or to the spinal cord.

The regulatory regions controlling expression of the sodium channel gene could be used in gene therapy to control expression of a therapeutic construct in cells expressing the channel.

Such regions would be isolated by using the cDNA as a probe to identify genomic clones carrying the gene and also flanking sequence e.g. cosmids. Fragments of the cosmids containing intron or flanking sequence would be used in a reporter gene assay in e.g. DRG cultures or transgenic animals and genomic fragments carrying e.g. promoter, enhancer or LCR activity identified.

The invention will now be further described with reference to the following examples:

Preferably the sodium channel is substantially free of material with which it is normally found in nature.

Preferably the sodium channel is in a substantially purified form.

Example 1 - Derivation of the sequence of a rat dorsal root ganglia (DRG) sodium channel cDNA by subtraction hybridisation methodology.

1.1 cDNA synthesis from DRG-derived poly-A+ RNA

Dorsal root ganglia (DRG) from all spinal levels of neonatal Sprague-Dawley male and female rats were frozen in liquid nitrogen. RNA is extracted using guanidine isothiocyanate and phenol/chloroform extraction (Chomczynski and Sacchi 1987 Anal Biochem 162,156-159).

Total RNA isolation - the nerve tissue is homogenised using a Polytron homogeniser in 1ml extraction buffer (23.6g guanidinium isothiocyanate, 5ml of 250 mM sodium citrate (pH 7.0) made up to 50ml with distilled water. To this is added 2.5ml 10% sarcosyl and 0.36ml ß-mercaptoethanol). 0.1ml of 2M sodium acetate (pH 4.0) is added followed by 1 ml phenol. After mixing, 0.2ml chloroform is added and this is shaken vigorously and placed on ice for 5 minutes. This is then centrifuged at 12,000 revolutions per minute (rpm) for 30 minutes at 4°C. The aqueous phase is transferred to a fresh tube, 1ml of isopropanol is added and this is left at -20°C for an hour followed by centrifuging at 12000 rpm for 30 minutes at 4°C. The pellet is dissolved in 0.1ml extraction buffer and is again extracted with isopropanol. The resulting pellet is washed with 70% ethanol and is resuspended in diethyl pyrocarbonate (DEPC)-treated water. 0.3M sodium acetate (pH5.2) and 2 volumes of ethanol are added and the mixture is placed at -20°C for 1 hour. The RNA is precipitated, washed again with 70% ethanol and resuspended in DEPC-treated water. The optical density is measured at 260 nanometres (nm) to calculate the yield of total RNA. Poly A+ RNA is isolated from the total RNA by oligo-dT cellulose chromatography (Aviv and Leder 1972 Proc Natl Acad Sci 69,1408-1411). The following procedures are carried out at 4°C as far as is possible. Oligo-dT cellulose (Sigma) is prepared by treatment with 0.1M sodium hydroxide for 5 minutes. The oligo-dT resin is poured into a column and is neutralised by washing with neutralising buffer (0.5 M potassium chloride, 0.01M Tris (Trizma base - Sigma - Tris(hydroxymethyl)aminomethane) (pH 7.5). The RNA solution is adjusted to 0.5M potassium chloride, 0.01M Tris (pH7.5) and is applied

to the top of the column. The first column eluate is re-applied to the column to ensure sticking of the mRNA to the oligo-dT in the column. The column is then washed with 70ml of neutralising buffer and the polyA+ RNA is eluted with 6ml 0.01M Tris (pH7.5) and 1ml fractions are collected. The poly A+ RNA is usually in fractions 2 to 5 and this is checked by measuring the optical density at 260nm. These fractions are pooled and ethanol precipitated overnight at -70°C, washed in 70% ethanol and then redissolved in deionised water at a concentration of 1mg/ml.

First strand cDNA was generated using 0.5µg DRG poly A+ mRNA, oligo-dT/Not-I primer adapters and SuperScript reverse transcriptase (Gibco-BRL) using methodology as described in example 2. One half of the cDNA was labelled by including 2 MBq ^{32}P dCTP (Amersham) in the reverse transcriptase reaction. Labelled cDNA is separated from unincorporated nucleotides on Nick columns (sephadex G50 - Pharmacia).

1.2 Enrichment of DRG-specific cDNA using subtraction hybridisation.

Poly A+ RNA from various tissues (10µg) is incubated with 10µg photoactivatable biotin (Clontech) in a total volume of 15µl and irradiated at 4°C for 30 minutes with a 250 watt sunlamp. The photobiotin is removed by extraction with butanol, and the cDNA co-precipitated with the biotinylated RNA without carrier RNA (Sive and St. John 1988 Nuc Ac Res 16,10937). Hybridisation is carried out at 58°C for 40 hours in 20% formamide, 50mM 3-(N-Morpholino)propane-sulphonic acid (MOPS) (pH 7.6), 0.2% sodium dodecyl sulphate (SDS), 0.5M sodium chloride, 5mM ethylenediaminetetraacetate (EDTA - Sigma). The total reaction volume is 5ml and the reaction is carried out under mineral oil, after an initial denaturation step of 2 minutes at 95°C. 100ml 50mM MOPS (pH 7.4), 0.5M sodium chloride, 5mM EDTA containing 20 units of streptavidin (BRL) is then added to the reaction mixture at room temperature, and the aqueous phase retained after two phenol /chloroform extraction steps. After sequential hybridisation with biotinylated mRNA from liver and kidney, followed by cortex and cerebellum, a 80-fold concentration of DRG-specific transcripts is

achieved.

One third of the 1-2 ng of residual cDNA is then G-tailed with terminal deoxynucleotide transferase at 37°C for 30 minutes. The polymerase chain reaction is used to amplify the cDNA using an oligo-dT-Not-I primer adapter and oligo-dC primers starting with the sequence AATTCCGA(C)^{10}. Amplification is carried out using 2 cycles of 95°C 1min, 45°C 1 min, 72°C 5min, followed by 2 cycles of 95°C for 1 minute , 58°C for 1 minute 72°C for 5 minutes. The resulting products are then separated on a 2% Nu-sieve agarose gel, and material running at a size of greater than 0.5 kilobase pairs (kb) is eluted and further amplified with 6 cycles of 45°C for 1 minute, 58°C for 1 minute and 72°C 5 for minutes. This material is further separated on a 2% Nu-sieve agarose gel, and the material running from 6kb on the gel is eluted and further amplified using the same PCR conditions for 27 cycles. The amplified DNA derived from this high molecular region is then further fractionated on a 2 % Nu-Sieve gel, and cDNA from 0.5 to 1.5kb, and from 1.5 to 5kb pooled.

1.3. Library Construction

10µg of the bacteriophage vector lambda-zap II (Stratagene) is restriction digested with NotI and EcoRI in high salt buffer overnight at 37°C followed by dephosphorylation using 1 unit of calf intestinal phosphatase (Promega) for 30 minutes at 37°C in 10mM Tris.HCl (pH9.5), 1mM spermidine, 0.1mM EDTA. DRG cDNA is digested with Klenow enzyme in the presence of dGTP and dCTP to construct an EcoRI site from the oligo-dC primer (see above) at the 5' end of the cDNA, and cut with NotI for directional cloning. The cDNA is ligated into the cloning vector bacteriophage lambda-zap II for 16 hours at 12°C. Recombinant phage DNA is then packaged into infective phage using Gigapack gold (Stratagene) and protocols specified by the suppliers. 0.1% of the packaged DNA is used to infect E.coli BB4 cells which are plated out to calculate the number of independent clones generated.

1.4 Differential Screening

The library is plated at a low density (10^3 clones/ 12 x 12

cm^2 dish) and screened using three sets of ^{32}P-labelled cDNA probes and multiple filter lifts. Replica filters are made by laying them onto the plated library plates, briefly drying them and then laying onto fresh agar plates to increase the quantity of phage and the subsequent hybridisation signals of lifts taken from them. The probes are derived from; a) cortex and cerebellum poly A+ RNA, b) DRG poly A+ RNA , and c) subtracted cDNA from DRG. The two mRNA probes are labelled with ^{32}P dCTP using a reaction mixture containing 2-5μg RNA, 50μl 5 x RT buffer, 25 μl 0.1M dithiothreitol (DTT), 12.5μl 10mM dATP, dGTP, dCTP, 30pM oligo-dT, 75 μl ^{32}P-dCTP (30MBq Amersham), 25μl 100uM dCTP, 2μl RNasin (2units/ml) and 2μl SuperScript reverse transcriptase (GibcoBRL) in a final volume of 250μl. The reaction is incubated at 39°C for 60 minutes, and the RNA subsequently destroyed by adding 250μl water, 55μl 1M NaOH, and incubating at 70°C for 20 minutes. The reaction mixture is neutralised with acidified Tris base (pH 2.0) and precipitated with carrier tRNA (Boehringer) with isopropanol. The subtracted and amplified double-stranded DRG cDNA is random-prime labelled with ^{32}P dATP (Gibco multiprime kit). Replica filters are then prehybridised for 4 hours at 68°C in hybridisation buffer. Hybridisation was carried out for 20 hours at 68°C in 4x SSC (20xSSC consists of 175.3g of sodium chloride and 88.2g of sodium citrate in 800ml of distilled water. The pH is adjusted to 7.0 with 10N sodium hydroxide and this is made to 1 litre with distilled water), 5x Denhardts solution containing 150 mg/ml salmon sperm DNA, 20mg/ml poly-U, 20mg/ml poly-C, 0.5% SDS (Sigma), 5mM EDTA. The filters are briefly washed in 2 x SSC at room temperature, then twice with 2 x SSC with 0.5% SDS at 68°C for 15 minutes, followed by a 20 minute wash in 0.5% SDS, 0.2 x SSC at 68°C. The filters are autoradiographed for up to 1 week on Kodak X-omat film. Plaques that hybridise with DRG probes but not cortex and cerebellum probes are picked, phage DNA prepared and the cloned inserts released for subcloning into pBluescript (Stratagene).

The positive plaques are picked by lining up the autoradiogram with the plate using orientation marks and taking a plug from the plate corresponding to the positive hybridisation signal. The phage is eluted from the plug in 0.5ml phage dilution buffer (10mM

Tris chloride (pH7.5) 10mM magnesium sulphate) and the phage re-infected into E.coli BB4 and replated at a density of 200 to 1000 plaques/150mm plate as a secondary purification step to ensure purity of the clones. The positive secondaries are then picked as described previously. In order to sub-clone the insert DNA from the positive recombinant phage, they need to be amplified. This is accomplished by plate lysis where the phage totally lyse the E.coli BB4. 0.2ml of phage suspension is mixed with 0.1ml of an overnight culture of E.coli. This is added to 2.5ml of top agar (16g bacto-tryptone 10^6 bacto-yeast extract 5g sodium chloride 7g bacto-agar in 900mls ditilled water) and plated onto $9cm^2$ agar plates. These are incubated overnight at 37°C. 5ml of phage dilution buffer is then added to the plates and is incubated overnight at 4°C or for 4 hours with gentle scraping at room temperature. The phage-containing buffer is then recovered, 0.1ml chloroform is added and this phage stock is titrated as above and stored at 4°C. Phage DNA is prepared by first infecting 1010 E.coli B44 with 109 plaque forming units (pfus) of phage in 3ml of phage dilution buffer and shaking at 37°C for 20 minutes. The infected bacteria are added to 400ml of L broth (1.6% bactotryptone, 0.5% (w/v) Bacto yeast extract, 0.5% (w/v) magnesium sulphate) with vigorous shaking at 37°C for 9 hours. When lysis has occurred, 10ml of chloroform is added and shaking is continued for a further 30 minutes. The culture is then cooled to room temperature and pancreatic RNAase and DNAase are added to 1ug/ml for 40 minutes. Sodium chloride is then added to 1M and is dissolved by swirling on ice. After centrifuging at 8000rpm for 10 minutes the supernatant is recovered. Polyethylene glycol (PEG 6000) is added to 10% w/v and is dissolved by stirring whilst on ice for 2 hours. After centrifuging for 8000rpm for 10 minutes at 4°C the pellet is resuspended in 8ml of phage dilution buffer. This is extracted with an equal volume of phenol/chloroform followed by purification on a caesium chloride gradient (0.675g/ml caesium chloride - 24 hours at 38000 rpm at 4°C). The opaque phage band is removed from the centrifugation tube and dialysed against 10mM sodium chloride, 50mM Tris (pH8.0), 10mM magnesium chloride for 2 hours. EDTA is then added to 20mM, proteinase K to 50ug/ml and SDS to 0.5% and is incubated at 65°C for 1 hour. After dialysis overnight

against TE pure phage DNA results. The cloned insert is digested from the purified phage DNA using restriction enzymes as previously described. Each phage insert is then ligated into a plasmid vector e.g. pBluescript - Clontech using a ligation reaction as previously described.

Clone characterisation.

The plasmids are cross hybridised with each other. Unique clones are further analysed by Northern blotting and sequencing. The clone/s showing transcript sizes and sequence comparable with sodium channels are then used as hybridisation probes to screen neonatal a rat DRG oligo dT-primed full length cDNA library to derive full length cDNA clones using methodology as described above and in example 2. Biological activity of the rat DRG sodium channel is confirmed as in examples 4 and 7 below.

Example 2 - Homology cloning of the human cDNA homologous to the rat DRG sodium channel cDNA (SNS-B).

2.1. Isolation of human ganglia total RNA

The starting material for the derivation of the human cDNA homologue of the rat DRG sodium channel cDNA is isolated human dorsal root ganglia or trigeminal ganglia or other cranial ganglia from post-mortem human material or foetuses. Total ribonucleic acid (RNA) is isolated from the human neural tissue by extraction in guanidinium isothiocyanate (Chomczynski and Sacchi 1987 Anal Biochem 162,156-159) as described in example 1.

2.2 Determination of the transcript size of the human homologue of the rat DRG sodium channel cDNA (SNS-B).

Human dorsal root ganglia total RNA is electrophoretically separated in a 1% (w/v) agarose gel containing a suitable denaturing agent e.g. formaldehyde (Lehrach et al 1977 Biochemistry 16,4743; Goldberg 1980 Proc Natl Acad Sci 77,5794; Seed 1982 in Genetic engineering: principles and methods (ed JK Setlow and A Hollaender) vol 4 p91

Plenum Publishing New York) or glyoxal/DMSO (McMaster GK and Carmichael GG 1977 Proc Natl Acad Sci 74,4835), followed by transfer of the RNA to a suitable membrane (e.g. nitrocellulose). The immobilised RNA is then hybridised to radioactive (or other suitable detection label) probes consisting of portions of the rat sodium channel cDNA sequence (see below). After washing of the membrane to remove non-hybridised probe, the hybridised probe is visualised using a suitable detection system (e.g. autoradiography for ^{32}P labelled probes) thus revealing the size of the human homologous mRNA molecule. Specifically, 20-30 µg total RNA from neonatal rat tissues are separated on 1.2% agarose -formaldehyde gels, and capillary blotted onto Hybond-N (Amersham) (Ninkina et al. 1993 Nuc Ac Res 21,3175-3182). The amounts of RNA on the blot are roughly equivalent, as judged by ethidium bromide staining of ribosomal RNA or by hybridisation with the ubiquitously expressed L-27 ribosomal protein transcripts (Le Beau et al. 1991 Nuc Ac Res 19,1337). Each Northern blot contains human DRG, cortex, cerebellum, liver kidney, spleen and heart RNA. Probes (50ng) are labelled with ^{32}P-dATP (Amersham) by random priming. Filters are prehybridised in 50% formaldehyde 5 x SSC containing 0.5% SDS, 5 x Denhardts solution (50x Denhardts contains 5g of Ficoll (Type 400, Pharmacia), 5g of polyvinylpyrrolidone, 5g of bovine serum albumin (Fraction V, Sigma) and water to 500ml), 100 mg/ml boiled salmon sperm DNA, 10 mg/ml poly-U and 10 mg/ml poly-C at 45°C for 6 hours. After 36 hours hybridisation in the same conditions, the filters are briefly washed in 2 x SSC at room temperature, then twice with 2 x SSC with 0.5% SDS at 68°C for 15 minutes, followed by a 20 minute wash in 0.5% SDS, 0.2 x SSC at 68°C. The filters are autoradiographed for up to 1 week on Kodak X-omat film. The transcript size is calculated from the signal from the gel in comparison with gel moleculr weight standard markers.

2.3 Production of a human DRG cDNA library

In order to produce a representative cDNA library from the human dorsal root ganglia messenger RNA (poly A+ mRNA) is first isolated from the total RNA pool using oligo-dT cellulose chromatography (Aviv and Leder 1972 Proc Natl Acad Sci 69,1408-1411)

using methodology described in example 1. Synthesis of the first strand of cDNA from the polyA+ RNA uses the enzyme RNA-dependent DNA polymerase (reverse transcriptase) to catalyse the reaction. The most commonly used method of second strand cDNA synthesis uses the product of first strand synthesis, a cDNA:mRNA hybrid, as a template for priming the second strand synthesis. (Gubler and Hoffman 1983 Gene 25,263)).

2.3.1. First strand cDNA synthesis - 20ug of human DRG polyA+ RNA is pre-treated to destroy secondary structure which may inhibit first strand cDNA synthesis. 20ug of polyA+ RNA, 1ul 1M Tris (pH7.5) are made up to a volume of 100ul with distilled water. This is incubated at 90°C for 2 minutes followed by cooling on ice. 4.8 ul of 100 mM methyl mercury is then added for 10 minutes at room temperature. 10ul of 0.7M ß-mercaptoethanol and 100 units of human placental RNAase inhibitor are then added for 5 minutes at room temperature. The first strand synthesis reaction consists of 8ul 20mM dATP, 5ul 20mM dCTP, 8ul 20mM dGTP 8ul 20mM dTTP, 10ul 1mg/ml oligo-dT (12-18), 20ul 1M Tris (pH 8.3) (at 45°C), 8ul 3M potassium chloride, 3.3ul 0.5M magnesium chloride, 3ul $a^{32}P$ dCTP, 100 units Superscript II reverse transcriptase (GibcoBRL) made up to 200ul with distilled water. This reaction mixture is incubated at 45°C for 45 minutes after which another 50 units of Superscript reverse transcriptase is added and incubated for a further 30 minutes at 45°C. EDTA is then added to 10mM to terminate the reaction and a phenol/chloroform extraction is carried out. The DNA is then precipitated using ammonium acetate (freezing in dry ice/ethanol before centrifuging), washed with 70% ethanol and resuspended in 50µl distilled water. The size of the single stranded DNA is assessed by electrophoretically separating it out on an agarose gel (1% w/v) and autortadiographing the result against markers.

2.3.2 Second strand synthesis - the second strand synthesis reaction mixture consists of 0.5µg human DRG single stranded DNA, 2µl 1M Tris (pH7.5), 1ul 0.5M magnesium chloride, 3.33µl 3M potassium chloride, 2µl 0.5M ammonium sulphate, 1.5µl 10mM ßnicotinamide adenine

dinucleotide (NAD), 4µl of each of the 1mM dNTPs, 5µl 1mg/ml bovine serum albumin (BSA), 1 unit RNAase-H, 25 units Klenow polymerase all made up to 100µl with distilled water. This is incubated at 12°C for 1 hour and then at 20°C for 1 hour. The reaction is stopped by addition of EDTA to 20mM followed by a phenol/chloroform extraction. The DNA is ethanol precipitated (-70°C overnight) and is then washed with 70% ethanol followed by resuspension in 20µl distilled water. Size is checked by gel electrophoresis and autoradiography.

2.3.3 Double stranded cDNA end repair - in order to add linkers to the end of the cDNA molecules for subsequent cloning, the ends must first be repaired. The human DRG cDNA is treated with 500 units/ml of S1 nuclease in 0.25M sodium chloride, 1mM zinc sulphate, 50mM sodium acetate (pH4.5). Incubation is at 30°C for 40 minutes followed by neutralisation with Tris (pH 8.0) to 0.2M. The DNA is again ethanol precipitated, washed in 70% ethanol and resuspended in 20ul distilled water. The size is again checked to ensure that S1 nuclease digestion has not radically reduced the average DNA fragment size. The repair reaction consists of 19µl cDNA, 3µl 10xT4 polymerase buffer (0.33M Tris acetate (pH7.9), 0.66M potassium acetate, 0.1M magnesium acetate, 1mg/ml BSA and 5mM DTT), 2µl of each dNTP at 2mM, 2µl T4 polymerase and 4µl distilled water. This is incubated at 37°C for 30 minutes followed by addition of 1µl Klenow polymerase for 1 hour at room temperature. The DNA is then ethanol precipitated, washed in 70% ethanol and resuspended in 5µl distilled water. In order to protect naturally occurring restriction sites within the cDNA from being cleaved, the cDNA is treated with a methylase before the addition of linkers. The reaction mixture consists of 5µl human DRG double stranded DNA, 1µl S- adenosylmethionine, 2µl 1mg/ml BSA, 2µl 5x methylase buffer (0.5M Tris (pH8.0), 5mM EDTA), 0.2µl EcoRI methylase (NEB). This is incubated at 37°C for 20 minutes followed by phenol extraction, ethanol precipitation washing with 70% ethanol and resuspension in 20µl distilled water.

2.3.4. Addition of linkers to cDNA - EcoRI linkers are ligated to the cDNA molecules to facilitate cloning into lambda vectors. The ligation

reaction mixture consists of 1µl 10x ligation buffer (0.5M Tris chloride (pH7.5), 0.1M magnesium chloride and 0.05M DTT), 1ul 10mM ATP, 100ng cDNA, 5µg EcoRI linkers, 1 unit T4 DNA ligase, distilled water to 10µl. The reaction is incubated at 37°C for 1 hour, followed by addition of 6 more units of T4 ligase and a further incubation overnight at 15°C. The ligated samples are ethanol precipitated, washed in 70% ethanol and resuspended in 10µl distilled water. The cDNA is then digested with EcoRI to cleave any linker concatamers formed in the ligation process. This restriction digestion reaction contains 10µl cDNA, 2µl high salt buffer (10mM magnesium chloride, 50mM Tris chloride (pH7.5), 1mM DTT, 100mM sodium chloride), 2µl EcoRI (10 units/ul - NEB) and distilled water to 20µl. The digestion is carried out for 3 hours. The ligation and digestion steps are monitored using gel elecrophoresis to monitor the size of the products.

2.3.5 Size fractionation of cDNA - in order to assure that the library is not swamped with short cDNA molecules and to remove linker molecules a column purification is carried out. A 1ml Sepharose 4B column is made in a 1 ml plastic pipette plugged with a small piece of glass wool. This is equilibrated with 0.1M sodium chloride in TE. The cDNA is loaded onto the column and 1 drop fractions are collected. 2µl aliquots of each fraction are analysed by gel electrophoresis and autoradiography to determine the sizes of the cDNA in each fraction. Fractions containing cDNA of about 800 base pairs and above are pooled and purified by ethanol precipitation and resuspending in 10µl distilled water.

2.3.6 Cloning of cDNA into bacteriophage vector

Bacteriophage vectors designed for the cloning and propagation of cDNA are provided ready-digested with EcoRI and with phosphatased ends from commercial sources (e.g. lambda gt10 from Stratagene). The prepared subtracted cDNA is ligated into lambda gt10 using a ligation rection consisting of ligase buffer and T4 DNA ligase (New England Biolabs) as described elsewhere in this document.

2.4 Labelling of cDNA fragments (probes) for library screening

The 3' untranslated region of the rat DRG sodium channel cDNA clone (SNS-B) is subcloned using appropriate restriction enzymes into a plasmid vector e.g. pBluescript - Stratagene. The cDNA insert which is to form the labelled probe is released from the vector via digestion with appropriate restriction enzymes and the insert is separated from the vector via electrophoresis in a 1% (w/v) agarose gel. After removal of the separated insert from the agarose gel and purification it is labelled by standard techniques such as random priming and polymerisation (Feinberg and Vogelstein 1983 Anal Biochem 132,6) or nick translation (Rigby et al 1977 J Mol Biol 113,237) with ^{32}P or DIG-labelled nucleotides. Alternatively, if the probe cDNA insert is cloned into a vector containing strong bacteriophage promoters to which DNA-dependant RNA polymerases bind (SP6, T3 or T7 polymerases), synthetic cRNA is produced by in vitro transcription which incorporates ^{32}P or digoxygenin nucleotides. Other regions of the rat DRG sodium channel cDNA can also be used as probes in a similar fashion for cDNA library screening or Northern blot analysis. Specifically, a probe is made using a kit such as the Pharmacia oligo labelling kit. This will radioactively label the rat DRG sodium channel cDNA fragment. 50ng of denatured DNA (place in boiling waterbath for 5 minutes), 3µl of ^{32}PdCTP (Amersham) and 10µl reagent mix is made up to 49µl with distilled water. 1µl of Klenow fragment is added and the mixture is incubated at 37°C for one hour. To remove unincorporated nucleotides, the reaction mixture is applied to a Nick column (Sephadex G50 - Pharmacia) followed by 400µl of TE (10mM Tris chloride (pH7.4) 1mM EDTA (pH8.0)). Another 400µl of TE is added and the eluate is collected. This contains the labelled DNA to be used as a hybridisation probe.

2.5 cDNA library screening

In order to detect recombinants containing human homologues of the rat DRG sodium channel the human DRG cDNA library is screened using moderate stringency hybridisation washes (50-60°C, 5 x SSC, 30 minutes), using radiolabelled or other labelled DNA or cRNA probes derived from the 3' untranslated region as described above. Libraries

are screened using standard methodologies involving the production of nitrocellulose or nylon membrane replicas of DNA from recombinant plaques formed on agar plates (Benton et al 1977 Science 196;180). These are then hybridised to single stranded nucleic acid probes (see above). Moderate stringency washes are carried out (see wash conditions for Northern analysis in section 2.2) Plaques which are positive on duplicate filters (i.e. not artefacts or background) are then purified by one or more rounds of replating after dilution to separate the colonies and further hybridisation screening. Resulting positive plaques are purified, DNA is extracted and the insert sizes of these clones is examined. The clones are cross-hybridised to each other using standard techniques (Sambrook et al 1989 Molecular Cloning Second Edition Cold Spring Harbour Laboratory Press) and distinct positive clones identified. Detailed protocols for cDNA library screening are given in example 1.

2.6 Derivation of a full-length clone of the human homologue of the rat DRG sodium channel cDNA.

Overlapping positive clones from above are identified by cross-hybridisation. They are then restriction mapped to identify their common portions and restriction fragments representing the separate portions from the overlapping clones are ligated together using standard cloning techniques (Sambrook et al 1989 Molecular Cloning Second Edition Cold Spring Harbour Laboratory Press). For example, the most 5' fragment will contain any 5' untranslated sequence, the start codon ATG and 5' coding sequence. The most 3' clone will contain the most 3' coding sequence, a stop codon and any 3' untranslated sequence, a poly A consensus sequence and possibly a poly A run. Thus a recombinant molecule is generated which contains the full cDNA sequence of the human homologue of the rat DRG sodium channel cDNA. If overlapping clones do not produce sufficient fragments to assemble a full length cDNA clone, the full length oligo dT-primed human DRG library is re-screened to isolate a full length clone. Alternatively, a full length clone is derived directly from the library screening.

2.7 Characterisation of the human homologue full-length clone

The cDNA sequence from the full-length clone is used as a probe in Northern blot analysis to detect the messenger RNA size in human tissue for comparison with the rat messenger RNA size (see sections 1.1 and 2.2 for methodology).

Confirmation of biological activity of the cloned cDNA is carried out via in vitro translation of the human sodium channel mRNA and it's expression in Xenopus oocytes in an analogous manner to that for the rat DRG-specific TTXi resistant sodium channel as described in examples 4 and 7.

cDNA sequences which are shown to have activity as defined above are completely sequenced using dideoxy-mediated chain termination sequencing protocols (Sanger et al 1977 Proc Natl Acad Sci 74,5463).

Example 3 - Polymerase chain reaction (PCR) approaches to clone the human DRG sodium channels using DNA sequence derived from the rat DRG sodium channel cDNA clone

Total RNA and poly A+ RNA is isolated from human dorsal root ganglia or trigeminal ganglia or other cranial ganglia from post-mortem human material or foetuses as described in example 2 above.

Random primers are hybridised to the RNA followed by polymerisation with MMLV reverse transcriptase to generate single stranded cDNA from the extracted human RNA.

Using degenerate PCR primers derived from relatively conserved regions of the known voltage-gated sodium channels (Figure 2), amplify the cDNA using the polymerase chain reaction (Saiki et al 1985 Science 230,1350). It is appreciated by those skilled in the art that there are many variables which can be manipulated in a PCR reaction to derive the homologous sequences required. These include but are not limited to varying cycle and step temperatures, cycle and step times, number of cycles, thermostable polymerase, Mg^{2+} concentration. It is also appreciated that greater specificity can be

gained by a second round of amplification utilising one or more nested primers derived
from further conserved sequence from the sodium channels.

Specifically, the above can be accomplished in the following manner. The first strand cDNA reaction consists of 1µg of total RNA made up to 13µl with DEPC-treated water and 1µl of 0.5µg/µl oligo(dT). This is heated to 70°c for 10 minutes and then incubated on ice for 1minute. The following is then added: 2µl of 10x synthesis buffer (200mM Tris chloride, 500mM potassium chloride, 25mM magnesium chloride, 1µg/µl BSA), 2µl of 0.1M DTT, 1µl of 200U/µl Superscript Reverse Transcriptase (Gibco BRL). This is incubated at room temperature for 10 minutes then at 42°C for 50 minutes. The reaction is then terminated by incubating for 15 minutes at 70°C. 1µl of E.coli RNase H (2U/µl) is added to the tube which is then incubated for 20 minutes at 37°C.

The PCR reaction is set up in a 0.5ml thin-walled Eppendorf tube. The following reagents are added:
10µl 10x PCR buffer, 1µl cDNA, 16µl dNTP's (25µl of 100µM of dATP, dCTP, dCTP and dGTP into 900µl sterile distilled water), 7µl of 25mM magnesium chloride, 1µl of Taq DNA polymerase (Amplitaq Perkin-Elmer) plus sterile distilled water to 94µl.

To each reaction tube a wax PCR bead is added (Perkin-Elmer) and the tube placed in a 70°C hot block for 1 minute. The tubes are allowed to cool until the wax sets and 3µl of each primer (33pM/µl) are added above the wax. The tubes are placed in a thermal cycler (Perkin-Elmer) and the following 3-step program used after an initial 94°C for 5 minutes; 92°C for 2 minutes, 55°C for 2 minutes, 72°C for 2 minutes for 35 cycles. A final polymerisation step is added at 72°C for 10 minutes. The reaction products are then run on a 1% agarose gel to assess the size of the products. In addition, control reactions are performed alongside the samples. These should be: 1/ all components without cDNA (negative control) and 2/ all reaction components with primers for constitutively expressed product e.g. §-actin or HPRT.

The products of the PCR reactions are examined on 0.8%-1.2% (w/v) agarose gels. Bands on the gel (visualised by staining with ethidium bromide and viewing under UV light) representing

amplification products of the approximate predicted size were then cut from the gel and the DNA purified. Further bands of interest are also identified by Southern blot analysis of the amplification products and probing of the resulting filters with labelled primers from further conserved regions e.g. those used for secondary amplification.

The resulting DNA is ligated into suitable vectors such as but not limited to pCR II (Invitrogen) or pGemT. Clones are then sequenced to identify those containing sequence with similarity to the rat DRG sodium channel sequence (SNS-B).

Clone analysis

Candidate clones from above are used to screen a human cDNA DRG library constructed using methods described in example 2. If a full length clone is not identified, positive overlapping clones which code for the full length human cDNA homologue are identified and a full length clone is then assembled as described in example 1. Biological activity is then confirmed as described in examples 4 and 7.

Example 4 - In vitro translation of rat and human DRG sodium channel in Xenopus laevis oocytes

In order to demonstrate the biological activity of the protein coded for by the rat DRG sodium channel cDNA sequence (SNS-B) and it's human homologue the complete double-stranded cDNA coding sequences are ligated into *in vitro* transcription vectors (including but not limited to the pGEM series, Promega) using one or more of the available restriction enzyme sites such that the cDNAs are inserted in the correct orientation. The constructs are then used to transform bacteria and constructs with the correct sequence in the correct orientation are identified via diagnostic restriction enzyme analysis and dideoxy-mediated chain termination DNA sequencing (Sanger et al 1977 Proc Natl Acad Sci 74,5463).

These constructs are then linearised at a restriction site downstream of the coding sequence and the linearised and purified plasmids are then utilised as a template for in vitro transcription. Sufficient quantities of synthetic mRNA are produced via in vitro

transcription of the cloned DNA using a DNA-dependent RNA polymerase from a bacteriophage that recognises a bacteriophage promoter found in the cloning vector. Examples of such polymerases include (but are not limited to) T3, T7 and SP6 RNA polymerase.

A variation on the above method is the synthesis of mRNA containing a 5' terminal cap structure (7-methylguanosine) to increase its stability and enhance its translation efficiency (Nielson and Shapiro 1986 Nuc Ac Res 14,5936). This is accomplished by the addition of 7-methylguanosine to the reaction mixture used for synthetic mRNA synthesis. The cap structure is incorporated into the 5' end of the transcripts as polymerisation occurs. Kits are available to facilitate this process e.g. mCAP RNA Capping Kit - Stratagene).

The synthetic RNA produced from the in vitro transcription is isolated and purified. It is then translated via microinjection into Xenopus laevis oocytes. 50nls of 1mg/ml synthetic RNA is micro-injected into stage 5 or stage 6 oocytes according to methods established in the literature (Gurdon et al (1983) Methods in Enzymol 101,370). After incubation to allow translation of the mRNAs the oocytes are analysed for expression of the DRG sodium channels via electrophysiological or other methods as described in example 7.

A further method for expression of functional sodium channels involves the nuclear injection of a Xenopus oocyte protein expression vector such as pOEV (Pfaff et al. 1990) which allows cloned DNA to be transcribed and translated directly in the oocyte. Since proteins translated in oocytes are post-translationally modified according to conserved eukaryotic signals, these cells offer a convenient system for performing structural and functional analyses of cloned genes. pOEV can be used for direct analysis of proteins encoded by cloned cDNAs without preparing mRNA in vitro, simplifying existing protocols for translating proteins in oocytes with a very high translational yield. Transcription of the vector in oocytes is driven by the promoter for the TPIIIA gene, which can generate 1-2 ng (per oocyte within 2 days) of stable mRNA template for translation. The vector also contains SP6 and T7 promoters for in vitro transcription to make mRNA and hybridization probes. DNA clones encoding SNS channel

transcripts are injected into oocyte nuclei and protein accumulated in the cell over a 2- to 10-day period. The presence of functional protein is then assessed using twin electrode voltage clamp as described in example 7.

Example 5 - Expression of rat and human DRG sodium channel in mammalian cells

In order to be able to establish a mammalian cell expression system capable of producing the sodium channel in a stable bioactive manner, constructs have to be first generated consisting of the cDNA of the channel in the correct vectors suitable for the cell system in which it is desired to express the protein. There are available a range of vectors containing strong promoters which drive expression in mammalian cells.

i/ Transient expression

In order to determine rapidly the bioactivity of a given cDNA it can be introduced directly into cells and resulting protein activity assayed 48-72 hours later. Although this does not result in a cell line which is stably expressing the protein of interest it does give a quick answer as to the biological activity of the molecule. Specifically, the cDNA representing the human or rat DRG sodium channel is ligated into appropriate vectors (including but not limited to pRc/RSV, pRc/CMV, pcDNA1 (Invitrogen)) using appropriate restriction enzymes such that the resulting construct contains the cDNA in the correct orientation and such that the heterologous promoter can drive expression of the transcription unit. The resulting expression constructs are introduced into appropriate cell lines including but not limited to COS-7 cells (an African Green Monkey Kidney cell line), HEK 293 cells (a human embryonic kidney cell line) and NIH3T3 cells (a murine fibroblastic cell line)). The DNA is introduced via standard methods (Sambrook et al 1989 Molecular Cloning Second Edition Cold Spring Harbour Laboratory Press) including but not limited to calcium phosphate transfection, electroporation or lipofectamine (Gibco) transfection. After the required incubation time at 37°C in a humidified incubator the cells are tested for the

presence of an active rat DRG sodium channel using methods described in example 7.

ii/Stable expression - the production of a stable expression system has several advantages over transient expression. A clonal cell line can be generated that a has a stable phenotype and in which the expression levels of the foreign protein can be characterised and , with some expression systems, controlled. Also, a range of vectors are available which incorporate genes coding for antibiotic resistance, thus allowing the selection of cells transfected with the constructs introduced. Cell lines of this type can be grown in tissue culture and can be frozen down for long-term storage. There are several systems available for accomplishing this e.g. CHO, CV-1, NIH-3T3.

Specifically COS-7 cells can be transfected by lipofection using Lipofectamine (GibcoBRL) in the manner. For each sample $2x10^6$ cells are seeded in a 90mm tissue culture plate the day prior to transfection. These are incubated overnight at 37°C in a CO_2 incubator to give 50-80% confluency the following day. The day of the transfection the following solutions are prepared in sterile 12 x 75mm tubes: Solution A: For each transfection, dilute 10-50µg of DNA into 990µl of serum-free media (Opti-MEM I Reduced Serum Medium GibccBRL). Solution B: For each transfection, dilute 50µl of Lipofectamine Reagent into 950µl serum-free medium. The two solutions are combined, mixed gently and incubated at room temp for 45 minutes. During this time the cells are rinsed once with serum-free medium. For each transfection 9ml of serum-free medium is added to the DNA-lipofectamine tubes. This solution is mixed gently and overlayed on the rinsed cells. The plates are incubated for 5 hours at 37°C in a CO_2 incubator. After the incubation the medium is replaced with fresh complete media and the cells returned to the incubator. Cells are assayed for activity 72 hours post transfection as detailed in examples 4 and 7. To ascertain the efficiency of transfection, ß-galactosidase in pcDNA3 is transfected alongside the DRG sodium channel cDNA. This control plate is stained for ß-galactosidase activity using a chromogenic substrate and the proportion of cells

staining calculated. For transient transfection of DRG the cDNA must first be cloned into a eucaryotic expression vector such as pcDNA3 (Invitrogen).

Example 6 - Expression of rat DRG sodium channel in insect cells

The baculovirus expression system uses baculovirus such as Autographa californica nuclear polyhedrosis virus (AcNPV) to produce large amounts of target protein in insect cells such as the Sf9 or 21 clonal cell lines derived from Spodoptera frugiperda cells. Expression of the highly abundant polyhedrin gene is non-essential in tissue culture and its strong promoter (polh) can be used for the synthesis of foreign gene products (Smith et al 1983 Mol Cell Biol 3,2156-2165). The polyhedrin promoter is maximally expressed very late in infection (20 hours post infection).

A transfer vector, where the rat DRG sodium channel cDNA is cloned downstream of the polh promoter, or another late promoter such as p10, is transfected into insect cells in conjunction with modified AcNPV viral DNA such as but not limited to BaculoGold DNA (PharMingen). The modified DNA contains a lethal mutation and is incapable of producing infectious viral particles after transfection. Co-transfection with a complementing transfer vector such as (but not limited to) pAcYM1 (Matsuura et al 1987 J Gen Virol 68,1233-1250) or pVL1392/3 (InVitrogen) allows the production of viable recombinant virus. Although more than 99% of the resultant virus particles should be derived from plasmid-rescued virus it is desirable to further purify the virus particles by plaque assay. To ensure that the recombinant stock is clonal, a single plaque is picked from the plaque assay and amplified to produce a recombinant viral stock. Once the recombinant phenotype is verified the viral stock can be used to infect insect cells and express functional rat DRG sodium channel. There are a number of variations in the methodology of baculovirus expression which may give increased expression (O'Reilly et al 1992 Baculovirus Expression Vectors: A Laboratory Manual. Oxford University Press). The expression of the rat or human DRG sodium channel is achieved by cloning of the cDNA into pVL1392 and introducing this into Sf21 insect cells.

Example 7 - Electrophysiological characterisation of cloned human and rat DRG sodium channel expression

Xenopus laevis oocytes are used tc express the channel after injection of the mRNA or cDNA in an expression vector. Expression would be transient and thus functional studies would be made at appropriate times after the injections. Comparison with mock-injected oocytes would demonstrate lack of the novel channel as an endogenously expressed characteristic. Standard two electrode voltage clamp (TEVC) techniques as described for example in Fraser, Moon & Djamgoz (1993) would be used to examine the characteristics of responses of ionic currents to changes in the applied membrane potential. Appropriately modified saline media would be used to manipulate the type of ionic currents detectable. The kinetics of activation and inactivation of the sodium current, its ionic selectivity, the effects of changes in ionic concentration of the extracellular medium on its reversal potential, and the sensitivity (or resistance) to TTX would be defining characteristics.

Similar electrophysiological studies would be undertaken to assess the success of functional expression in a permanently or transiently expressing mammalian cell line, but patch clamp methods would be more suitable than TEVC. Whole cell, cell-attached patch, inside-out patch or outside-out patch configurations as described for example by Hamill et al. (1981) and Fenwick et al. (1982) might be used to assess the channel characteristics.

Hamill, O.P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F.J. (1981) Pfl gers Arch. 391: 85-100.

Fenwick, E.M., Marty, A. & Neher.E. (1982) J. Physiol. 331 599-635 Fraser, S.P., Moon, C. & Djamgoz, M.B.A. (1993) Electrophysiology of Xenopus oocytes: an expression system in molecular neurobiology. In: Electrophysiology: A practical approach. Wallis, D.I., ed. Oxford University Press. Chapter 4 pp. 65-86.

For example, Isolated transfected cells (see above) will be voltage-clamped using the whole-cell variant of the patch clamp technique for recording the expressed sodium channel

current.

Recordings will be obtained at room temperature (22-24 oC). Both external and internal recording solutions will be used to isolate Na+ currents as previously described (Lalik et al., Am. J. Physiol. 264:C803-C809, 1992; West et al., Neuron 8:59-70, 1992). External solution (mM): sodium chloride, 65; choline chloride, 50; TEA-Cl, 20, KCl, 1.5; calcium chloride, 1; magnesium chloride, 5; glucose 5; HEPES, 5; at a pH 7.4 and and osmolality of 320. Internal solution (mM):CsF, 90; CsCl, 60; sodium chloride, 10; MgCl, 2;EGTA, 10; HEPES, 10 at pH 7.2 andan osmolality of 315.

The kinetics and voltage parameters of the expressed sodium channel current will be examined and campared with data existing in the literature. These include current-voltage relationships and peak current amplitude. Cells will be voltage-clamped at -70 mV and depolarizing pulses to 50 mV (at 10 mV increments) will be used to generate currents.

The pharmacology of the expressed sodium channel current will be examined with the Na channel blocker, tetrodotoxin (TTX). To date sodium channels have been classified as TTX-sensitive and TTX-resistant: block by low (1-30 nM) and high (> 1 uM) concentrations of TXX, respectively (Elliot & Elliot, J. Physiol. (Lond.) 463:39-56, 1993; Yang et al., J. Neurosci. 12:268-277, 1992; W1992).

The channel is unaffected by concentrations lower than 1 micromolar tetrodotoxin, and is only partially blocked by concentrations as high as 10 micromolar tetrodotoxin.

Example 8 - Production of purified channel

Using a commercial coupled transcription-translation system, 35-S methionine labelled protein products of the SNS clone can be generated (see Figure 3). The size of the resulting protein when assessed by SDS-polyacrylamide gel electrophoresis confirms the predicted size of the protein deduced by DNA sequencing. The system used is the Promega TNT system (Promega Technical Bulletin 126 1993). The experiment is carried out precisely according to the protocol provided (see Figure 3).

Example 9 - Use of rat or human sodium channel in screening assays

Cell lines expressing the cloned sodium channels could be used to determine the effects of drugs on the ability of the channels to pass sodium ions across the cell membranes, e.g to block the channels or to enhance their opening. Since the channel activation is voltage dependent, depolarising conditions are likely to be required for observation of baseline activity that would be modified by drug actions. Depolarisation could be achieved by for example raising extracellular potassium ion concentration to 20 or 40 mM, or by repeated electrical pulses applied to the bathing medium. Detection of the activation of sodium conducting channels could be achieved by flux of radiolabelled sodium ions, guanidine or by reporter gene activation leading to for example a colour change or to fluorescence of a light emitting protein. Subsequent confirmation of the effectiveness of the drug action on sodium channel activity would require electrophysiological studies similar to those described above.

Example 10 In vitro influx assays

1. 22Na+ influx assay: A modified assay has been adapted from methods reported by Tamkum and Catterall (Mol Pharm. 19:78, 1981). Oocytes or cells expressing the sodium channel gene is suspended in a buffer containing 0.13 M sodium chloride, 5 mM KCl, 0.8 mM $MgSO_4$, 50 mM HEPES-Tris (pH 7.4), and 5.5 mM glucose. Aliquots of the cell suspension are added a buffer containing 22NaCl (1.3 uCi/ml, New England Nuclear, Boston, MA), 0.128 M choline chloride, 2.66 mM sodium chloride, 5.4 mM KCl, 0.8 mM $MgSO_4$, 50 mM HEPES-Tris (pH 7.4), 5 mM ouabain 1mg/ml bovine serum albumin, and 5.5 mM glucose and then incubated at 37 oC for 20 sec in either the presence or absence of 100 uM veratridine (Sigma Chemical Co., St Louis, MO). The influx assay is stopped by the addition of 3 ml of ice-cold wash buffer containing 0.163 M sodium chloride, 0.8 mM $MgSO_4$, 1.8 mM $CaCl_2$, 50 mM HEPES-Tris (pH 7.4) and 1mg/ml bovine serum albumin, collected on a glass fiber filter (Whatman GF/C), and washed twices with 3 ml of wash buffer. Radioactive incorporation is determined by with a gammacounter. The

specific tetrodotoxin-resistant influx is measured by the difference in 22Na+ uptake in the absence or the presence of 10 uM transmethrin or 1 uM (+) trans allethrin. The tetrodotoxin-sensitive influx is measured by the difference in 22Na+ uptake in the absence or the presence of 1 uM tetrodotoxin (Sigma Chemical Co., St Louis, MO).

Guanidine influx: Another assay is modified from the method described by Reith (Eur. J. Pharmacol. 188:33, 1990). In this assay sodium ions are substituted with guanidinium ions. Oocytes or cells are washed twice with a buffer containing 4.74 mM KCl, 1.25 mM $CaCl_2$, 1.2 mM KH2PO4, 1.18 mM $MgSO_4$, 22 mM HEPES (pH 7.2), 22 mM choline chloride and 11 mM glucose. The oocytes or cells are suspended in the same buffer containing 250 uM guanidine for 5 min at 19-25 oC. An aliquot of 14C-labelled guanidine hydrochloride (30-50 mCi/mmol supplied by New England Nuclear, Boston, MA) is added in the absence or presence of 10 uM veratridine, and the mixture is incubated for 3 min. The uptake reaction is stopped by filtration through Whatman GF/F filters and followed by 2 5 ml washes with ice-cold 0.9% saline. Radioactive incorporation is determined by scintillation counting.

Example 11

In order to measure the expression of sodium channels in in vitro systems, as well as to analyse distribution and relative level of expression in vivo, and to attempt to block function, polyclonal and monoclonal antibodies will be generated to peptide and protein fragments derived from SNS protein sequence shown in Figure 1.

a) Immunogens

Glutathione-sulphotransferase (GST) - fusion proteins will be constructed (Smith and Johnson 1988) using PGEX vectors obtained from Pharmacia. Fusion proteins including both intracellular and extracellular loops with little homology with known sodium channels other than SNS-B will be produced. One such method involves subcloning of fragments into pGex-5X3 or pGEX 4t-2 to produce

in-frame fusion proteins encoding extracellular, intracellular or C-terminal domains as shown in detailed maps in Figure 4. The pGEX fusion vectors are transformed into E. coli XL-1 blue cells or other appropriate cells grown in the presence of ampicillin. After the cultures have reached an optical density of OD600 > 0.5, fusion protein synthesis is induced by the addition of 100 micromolar IPTG, and the cultures further incubated for 1- 4 hours. The cells are harvested by centrifugation and washed in ice cold phosphate buffered saline. The resuting pellet (dissolved in 300 microlitres PBS from each 50 ml culture) is then sonicated on ice using a 2mm diameter probe, and the lysed cells microfuged to remove debris. 50 microlitres of glutathione-agarose beads are then added to each pellet, and after gentle mixing for 2 minutes at room temperature, the beads are washed by successive spins in PBS. The washed beads are then boiled in Laemmli gel sample buffer, and applied to 10% polyacrylamide SDS gels. Material migrating at the rpredicted molecular weight is identified on the gel by brief staining with coommassie blue, and comparison with molecular weight markers. This material is then electroeluted from the gel and used as an immunogen as described below.

b) Antibody production

Female Balb/c mice are immunised intraperiteonally with 1-100 micrograms of GST fusion protein emulisfied in Freunds complete adjuvant. After 4 weeks, the animals will be further immmnised with fusion proteins (1-100 micrograms) emulsified in Freunds incomplete adjuvant. Four weeks later, the animals will be immunised intraperitoneally with a further 1-100 micrograms of GST fusion protein emulsified with Freunds incomplete adjuvant. Seven days later, the animals will be tail bled, and their serum assessed for the production of antibodies to the immunogen by the following screen; (protocols for the production of rabbit polyclonal serum are the same, except that all injections are subcutaneous, and 10 times as much immunogen is used. Polyclonal rabbit serum are isolated from ear-vein bleeds.)

Serial ten-fold dilutions of the sera (1;100 to 1; 1000,000) in phosphate buffered saline (PBS) containing 0.5% NP-40 and 1% normal goat serum will be applied to 4% paraformaldehyde-fixed 10 micron sections of neonatal rat spinal cord previously treated with 10% goat serum in PBS, After overnight incubation, the sections are washed in PBS, and further incubated in the dark with 1;200 FITC-conjugated F(ab)2 fragment of goat anti-mouse antibodies for 2 hours in PBS containing 1% normal goat serum. The sections are further washed in PBS, mounted in Citifluor, and examined by fluorescence microscopy. Those sera that show specific staining of laminar II in the spinal cord will be retained, and the mice generating such antibodies subsequently used for the production of monoclonal antibodies.
Three weeks later, mice producing useful antibodies are immunised with GST-fusion proteins without adjuvant. After 3 days, the animals are killed, their spleens removed, and the lymphocytes fused with the thymidine kinase-negative myeloma line NSC, using polyethylene glycol. The fused cells from each experiment are grown up in 3 x 24 well plates in the presence of DMEM medium containing 10% fotal calf serum and hypoxanthine, aminopterin and thymidine (HAT) medium to kill the myeloma cells (Kohler and Milstein 1976). The tissue culture supernatants from wells containing hybridomas are further screened by immunofluorescence as described above, and cells from positive wells cloned by limiting dilution. Antibody from the positive testing cloned hybridomas is then used to Western blot extracts of rat dorsal root ganglia, to detemine if the antibody recognises a band of size approximately 200,000, confirming the specificity of the monoclonal antibody for the SNS sodium channel. Those antibodies directed against extracellular domains that test positive by both of these criteria will then be assessed for function blocking activity in electrophysiological tests of sodium channel function (see example 7), and in screens relying on ion flux or dye-based assays in cells lines expressing sodium channel (see examples 9 and 10).

Example 12 - Cell-type distribution of expression

In situ hybridisation demonstrates the presence of SNS in a subset of sensory neurons. An SNS fragment between positions 1740 and 1960 was sub-cloned into pGem4z, and DIG-UTP labelled sense or antisense cRNA generated. Sample preparation, hybridisation, and visualisation of in situ hybridisation with alkaline phosphatase conjugated anti-DIG antibodies was carried out exactly as described in (Schaeren-Wimers N. and Gerfin-Moser A. 1993).

References

Catterall W .A. (1992) Physiol. Rev. 72, S4-S47.
Cohen S.A. and Barchi R.L. (1993) Int. Rev. Cytology 137c, 55-103.
Hodgkin A.L. and Huxley A.F. (1952) J. Physiol. 116, 473-496.
Hille B. (1991) Ionic channels in excitable membranes (Sinauer Sunderland MA)
Jeftjina S. (1994) Brain Res. 639, 125-134.
Kohler G. and Milstein C. (1976) Eur J. Immunol 6, 511-519
Lewin B. (1995) Genes V Oxford University Press, Oxford.
Melton D. et al. (1984) Nucleic Acids Res. 12, 7035
Nowycky M. (1993) in Sensory Neurons (Ed Scott S.) OUP, Oxford.
Omri G. and Meir H. (1990) J. Membrane Biol. 115, 13-29
Pearce R.J. and Duchen M.R. (1994) Neuroscience 63, 1041-1056
Pfaff SL; Tamkun-MM; Taylor-WL (1990 Anal-Biochem. 1990 188 192-195
Schaeren-Wimers N. and Gerfin-Moser A. (1993) Histochemistry 100, 431-440.
Smith D.B. and Johnson K.S. (1988) Gene 67, 31-40.

In the accompanying Figures:

Figure 1a shows the nucleic acid and amino acid sequences of the sodium channel specific to the rat DRG (SNS-B).

Figure 1b shows the nucleic acid and amino acid sequences of a splice variant sodium channel specific to the rat DRG (SNS-B).

Figure 1c shows the nucleic acid and amino acid sequences of a short form splice variant sodium channel specific to the rat DRG

(SNS-B).

Figure 1d shows the structure of the SNS-B voltage-gated sodium channel in pGEM-3Z.

Figure 1e shows a schematised drawing of a known voltage-gated sodium channel.

Figure 2 shows sequences of examples of PCR primers for isolation of human clone probes.

Figure 3 shows a film of ^{35}S radio-labelled SNS-B voltage-gated sodium channel protein in a coupled transcription/translation system.

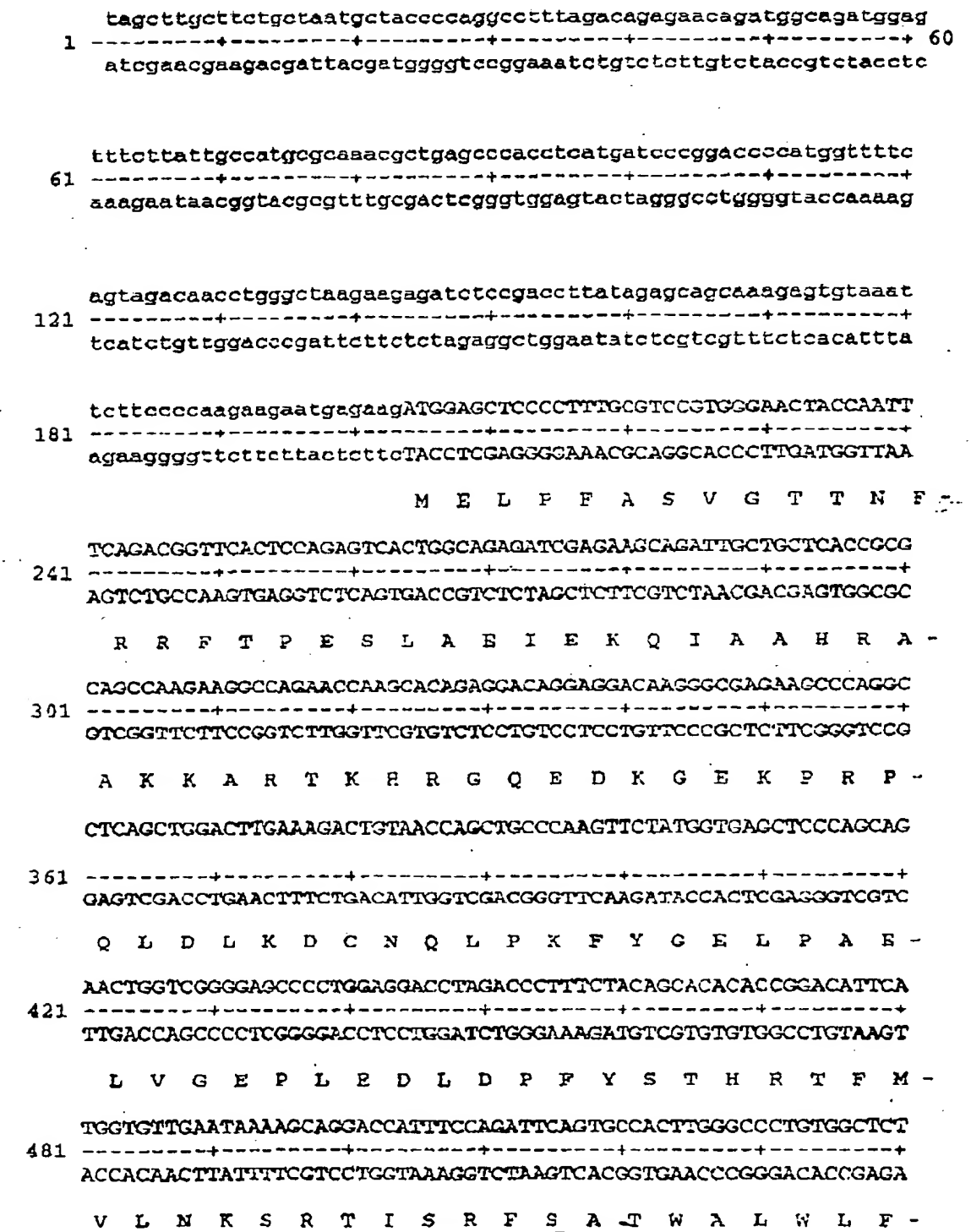
Figure 4a ans **Figure 4b** show SNS-GST fusion protein constructs for antibody generation.

RP09541

RMT/KEB : 28JUN95

Figure 1a

Nucleic acid and amino acid sequence of TTX1 DRG sodium channel



```
    TCAGTCCCTTCAACCTGATCAGAAGAACAGCCATCAAAGTGTCTGTCCATTCCTGGTCCT
541 ---------+---------+---------+---------+---------+---------+
    AGTCAGGGAAGTTGGACTAGTCTTCTTGTCGGTAGTTTCACAGACAGGTAAGGACCAAGA

      S  P  F  N  L  I  R  R  T  A  I  K  V  S  V  H  S  W  F  S -


    CCATATTCATCACCATCACTATTTTGGTCAACTGCGTGTGCATGACCCGAACTGATCTTC
601 ---------+---------+---------+---------+---------+---------+
    GGTATAAGTAGTGGTAGTGATAAAACCAGTTGACGCACACGTACTGGGCTTGACTAGAAG

      I  F  I  T  I  T  I  L  V  N  C  V  C  M  T  R  T  D  L  P -

    CAGAGAAAGTCGAGTACGTCTTCACTGTCATTTACACCTTCGAGGCTCTGATTAAGATAC
661 ---------+---------+---------+---------+---------+---------+
    GTCTCTTTCAGCTCATGCAGAAGTGACAGTAAATGTGGAAGCTCCGAGACTAATTCTATG

      E  K  V  E  Y  V  F  T  V  I  Y  T  F  E  A  L  I  K  I  L -

    TGGCAAGAGGGTTTTGTCTAAATGAGTTCACTTATCTTCGAGATCCGTGGAACTGGCTGG
721 ---------+---------+---------+---------+---------+---------+
    ACCGTTCTCCCAAAACAGATTTACTCAAGTGAATAGAAGCTCTAGGCACCTTGACCGACC

      A  R  G  F  C  L  N  E  F  T  Y  L  R  D  P  W  N  W  L  D -

    ACTTCAGTGTCATTACCTTGGCGTATGTGGGTGCAGCGATAGACCTCCGAGGAATCTCAG
781 ---------+---------+---------+---------+---------+---------+
    TGAAGTCACAGTAATGGAACCGCATACACCCACGTCGCTATCTGGAGGCTCCTTAGAGTC

      F  S  V  I  T  L  A  Y  V  G  A  A  I  D  L  R  G  I  S  G -

    GCCTGCGGACATTCCGAGttctcagagccctgaaaactgtttctgtgatcccaggactga
841 ---------+---------+---------+---------+---------+---------+
    CGGACGCCTGTAAGGCTCaagagtctcgggacttttgacaaagacactagggtcctgact

      L  R  T  F  R  V  L  R  A  L  K  T  V  S  V  I  P  G  L  K -

    aggtcatcgtgggagccctgatccactcagtgaggaagctggccgacgtgactatcctca
901 ---------+---------+---------+---------+---------+---------+
    tccagtagcaccctcgggactaggtgagtcactccttcgaccggctgcactgataggagt

      V  I  V  G  A  L  I  H  S  V  R  K  L  A  D  V  T  I  L  T -

    cagtcttctgcctgagcgtcttcgccttggtgggcctgcagctctttaaggggaaccttа
961 ---------+---------+---------+---------+---------+---------+
    gtcagaagacggactcgcagaagcggaaccacccggacgtcgagaaattccccttggaat

      V  F  C  L  S  V  F  A  L  V  G  L  Q  L  F  K  G  N  L  K -

    agaacaaatgcatcaggaacggaacagatccccacaaggccgacaacctctcatctgaaa
1021---------+---------+---------+---------+---------+---------+
    tcttgtttacgtagtccttgccttgtctaggggtgttccgactgttggagagtagacttt

      N  K  C  I  R  N  G  T  D  P  H  K  A  D  N  L  S  S  E  N -

    tggcagaatacgtctccatcaagccatggtactacggatcccttactgtgcggcaatgggt
1081---------+---------+---------+---------+---------+---------+
    accgtcttatgcagaggtagttcggaccatgatgcctagggaatgacacgccgttaccca

      A  E  Y  V  S  I  K  P  G  T  T  D  P  L  L  C  G  N  G  S -
```

```
    ctgatgctggtcactgccctggaggctatgtctgcctgaaaactcctgacaacccggatt
1141 ---------+---------+---------+---------+---------+---------+
    gactacgaccagtgacgggacctccgatacagacggacttttgaggactgttgggcctaa

     D  A  G  H  C  P  G  G  Y  V  C  L  K  T  P  D  N  P  D  F -


    ttaactacaccagctttgattcctttgcgtgggcattcctctcactgttccgcctcatga
1201 ---------+---------+---------+---------+---------+---------+
    aattgatgtggtcgaaactaaggaaacgcacccgtaaggagagtgacaaggcggagtact

     N  Y  T  S  F  D  S  F  A  W  A  F  L  S  L  F  R  L  M  T -


    cgcaggactcctgggagcgcctgtaccagcagacactccgggcttctgggaaaatgtaca
1261 ---------+---------+---------+---------+---------+---------+
    gcgtcctgaggaccctcgcggacatggtcgtctgtgaggcccgaagacccttttacatgt

      Q  D  S  W  E  R  L  Y  Q  Q  T  L  R  A  S  G  K  M  Y  M -


    tggtcttttttcgtgctggttatttccttggatcgttctacctggtcaatttgatcttgg
1321 ---------+---------+---------+---------+---------+---------+
    accagaaaaagcacgaccaataaaaggaacctagcaagatggaccagttaaaactagaacc

      V  F  F  V  L  V  I  F  L  G  S  F  Y  L  V  N  L  I  L  A -


    ccgtggtcaccatggcgtatgaagagcagagccaggcaacaattgcagaaatcgaagcca
1381 ---------+---------+---------+---------+---------+---------+
    ggcaccagtggtaccgcatacttctcgtctcggtccgttgttaacgtctttagcttcggt

      V  V  T  M  A  Y  E  E  Q  S  Q  A  T  I  A  E  I  E  A  K -


    aggaaaaaaagttccaggaagcccttgaggtgctgcagaaggaacaggaggtgctggcag
1441 ---------+---------+---------+---------+---------+---------+
    tcctttttttcaaggtccttcgggaactccacgacgtcttccttgtcctccacgaccgtc

      E  K  K  F  Q  E  A  L  E  V  L  Q  K  E  Q  E  V  L  A  A -


    ccctggggattgacacgacctcgctccagtcccacagtggatcacccttagcctccaaaa
1501 ---------+---------+---------+---------+---------+---------+
    gggacccctaactgtgctggagcgaggtcagggtgtcacctagtgggaatcggaggtttt

      L  G  I  D  T  T  S  L  Q  S  H  S  G  S  P  L  A  S  K  N -


    acgccaatgagagaagacccagggtgaaatcaagggtgtcagagggctccacggatgaca
1561 ---------+---------+---------+---------+---------+---------+
    tgcggttactctcttctgggtcccactttagttcccacagtctcccgaggtgcctactgt

      A  N  E  R  R  P  R  V  K  S  R  V  S  E  G  S  T  D  D  N -


    acaggtcaccccaatctgacccttacaaccagcgcaggatgtctttcctaggcctgtctt
1621 ---------+---------+---------+---------+---------+---------+
    tgtccagtggggttagactgggaatgttggtcgcgtcctacagaaaggatccggacagaa

      R  S  P  Q  S  D  P  Y  N  Q  R  R  M  S  F  L  G  L  S  S -


    caggaagacgcagggctagccacggcagtgtgttccacttccgagcgcccagccaagaca
1681 ---------+---------+---------+---------+---------+---------+
    gtccttctgcgtcccgatcggtgccgtcacacaaggtgaaggctcgcgggtcggttctgt

      Q  R  R  R  A  S  H  G  S  V  F  H  F  R  A  P  S  Q  D  I -
```

```
      tctcatttcctgacgggatcaccccctgatgatggggtctttcacggagaccaggaaagcc
1741  ---------+---------+---------+---------+---------+---------+
      agagtaaaggactgccctagtggggactactaccccagaaagtgcctctggtcctttcgg

       S  F  P  D  G  I  T  P  D  D  G  V  F  H  G  D  Q  E  S  R -

      gtcgaggttccatattgctgggcaggggtgctgggcagacaggtccactccccaggagcc
1801  ---------+---------+---------+---------+---------+---------+
      cagctccaaggtataacgacccgtccccacgacccgtctgtccaggtgaggggtcctcgg

       R  G  S  I  L  L  G  R  G  A  G  Q  T  G  P  L  P  R  S  P -

      cactgcctcagtcccccaaccctggccgtagacatggagaagagggacagctcggagtgc
1861  ---------+---------+---------+---------+---------+---------+
      gtgacggagtcaggggggttgggaccggcatctgtacctcttctccctgtcgagcctcacg

        L  P  Q  S  P  N  P  G  R  R  H  G  E  E  G  Q  L  G  V  P -

      ccactggtgagcttaccgctggagcgcctgaaggccccggcactgcacactacagggcaga
1921  ---------+---------+---------+---------+---------+---------+
      ggtgaccactcgaatggcgacctcgcggacttccgggccgtgacgtgtgatgtcccgtct

       T  G  E  L  T  A  G  A  P  E  G  P  A  L  H  T  T  G  Q  K -

      agagcttcctgtctgcgggctacttgaacgaacctttccgagcacagagggccatgagcg
1981  ---------+---------+---------+---------+---------+---------+
      tctcgaaggacagacgcccgatgaacttgcttggaaaggctcgtgtctccccggtactcgc

       S  F  L  S  A  G  Y  L  N  E  P  F  R  A  Q  R  A  M  S  V -

      ttgtcagtatcatgacttctgtcattgaggagcttgaagagtctaagctgaagtgcccac
2041  ---------+---------+---------+---------+---------+---------+
      aacagtcatagtactgaagacagtaactcctcgaacttctcagattcgacttcacgggtg

        V  S  I  M  T  S  V  I  E  E  L  E  E  S  K  L  K  C  P  P -

      cctgcttgatcagcttcgctcagaagtatctgatctgggagtgctgccccaagtggagga
2101  ---------+---------+---------+---------+---------+---------+
      ggacgaactagtcgaagcgagtcttcatagactagaccctcacgacggggttcacctcct

       C  L  I  S  F  A  Q  K  Y  L  I  W  E  C  C  P  K  W  R  K -

      agttcaagatggcgctgttcgagctggtgactgacccttcgcagagcttaccatcaccc
2161  ---------+---------+---------+---------+---------+---------+
      tcaagttctaccgcgacaagctcgaccactgactggggaagcgtctcgaatggtagtggg

       F  K  M  A  L  F  E  L  V  T  D  P  F  A  E  L  T  I  T  L -

      tctgcatcgtggtgaacaccgtcttcatggccatggagcactaccccatgaccgatgcct
2221  ---------+---------+---------+---------+---------+---------+
      agacgtagcaccacttgtggcagaagtaccggtacctcgtgatggggtactggctacgga

       C  I  V  V  N  T  V  F  M  A  M  E  H  Y  P  M  T  D  A  P -

      tcgatgccatgcttcaagccggcaacattgtcttcaccgtgtttttcacaatggagatgg
2281  ---------+---------+---------+---------+---------+---------+
      agctacggtacgaagttcggccgttgtaacagaagtggcacaaaaagtgttacctctacc

       D  A  M  L  Q  A  G  N  I  V  F  T  V  F  F  T  M  E  M  A -
```

```
    ccttcaagatcattgccttcgacccctactattacttccagaagaagtggaatatcttcg
2341 ---------+---------+---------+---------+---------+---------+
    ggaagttctagtaacggaagctggggatgataatgaaggtcttcttcaccttatagaagc

     F  K  I  I  A  F  D  P  Y  Y  Y  F  Q  K  K  W  N  I  F  D -

    actgtgtcatcgtcaccgtgagccttctggagctgagtgcatccaagaagggcagcctgt
2401 ---------+---------+---------+---------+---------+---------+
    tgacacagtagcagtggcactcggaagacctcgactcacgtaggttcttcccgtcggaca

     C  V  I  V  T  V  S  L  L  E  L  S  A  S  K  K  G  S  L  S -

    ctgtgctccgtaccttacgcttgctgcgggtcttcaagctggccaagtcctggcccaccc
2461 ---------+---------+---------+---------+---------+---------+
    gacacgaggcatggaatgcgaacgacgcccagaagttcgaccggttcaggaccgggtggg

     V  L  R  T  L  R  L  L  R  V  F  K  L  A  K  S  W  P  T  L -

    tgaacaccctcatcaagatcatcgggaactcagtgggggccctcggcaacctgaccttta
2521 ---------+---------+---------+---------+---------+---------+
    acttgtgggagtagttctagtagcccttgagtcacccccgggacccgttggactggaaat

     N  T  L  I  K  I  I  G  N  S  V  G  A  L  G  N  L  T  F  I -

    tcctggccatcatcgtcttcatcttcgccctggtcggaaagcagcttctctcagaggact
2581 ---------+---------+---------+---------+---------+---------+
    aggaccggtagtagcagaagtagaagcgggaccagcctttcgtcgaagagagtctcctga

     L  A  I  I  V  F  I  F  A  L  V  G  K  Q  L  L  S  E  D  Y -

    acgggtgccgcaaggacggcgtctccgtgtggaacggcgagaagctccgctggcacatgt
2641 ---------+---------+---------+---------+---------+---------+
    tgcccacggcgttcctgccgcagaggcacaccttgccgctcttcgaggcgaccgtgtaca

      G  C  R  K  D  G  V  S  V  W  N  G  E  K  L  R  W  H  M  C -

    gtgacttcttccattccttcctggtcgtcttccgaatcctctgcggggagtggatcgaga
2701 ---------+---------+---------+---------+---------+---------+
    cactgaagaaggtaaggaaggaccagcagaaggcttaggagacgccccctcacctagctct

     D  F  F  H  S  F  L  V  V  F  R  I  L  C  G  E  W  I  E  N -

    acatgtgggtctgcatggaggtcagccagaaatccatctgcctcatcctcttcttgactg
2761 ---------+---------+---------+---------+---------+---------+
    tgtacacccagacgtacctccagtcggtctttaggtagacggagtaggagaagaactgac

     M  W  V  C  M  E  V  S  Q  K  S  I  C  L  I  L  F  L  T  V -

    tgatggtgctgggcaacctagtggtgctcaacctttcatcgctttactgctgaactcct
2821 ---------+---------+---------+---------+---------+---------+
    actaccacgacccgttggatcaccacgagttggaaaagtagcgaaatgacgacttgagga

      M  V  L  G  N  L  V  V  L  N  L  F  I  A  L  L  L  N  S  F -

    tcagcgcggacaacctcacggctccagaggatgacggggaggtgaacaacttgcagttag
2881 ---------+---------+---------+---------+---------+---------+
    agtcgcgcctgttggagtgccgaggtcctcctactgccccctccacttgttgaacgtcaatc

     S  A  D  N  L  T  A  P  E  D  D  G  E  V  N  N  L  Q  L  A -

    cactggccaggatccaggtacttggccatcgggccagcagggccagcgccagttacatca
```

```
2941 ---------+---------+---------+---------+---------+---------+
     gtgaccggtcctaggtccatgaaccggtagcccggtcgtcccggtcgcggtcaatgtagt

       L  A  R  I  Q  V  L  G  H  R  A  S  R  A  S  A  S  Y  I  S -

     gcagccactgccgattccactggcccaaggtggagacccagctgggcatgaagcccccac

3001 ---------+---------+---------+---------+---------+---------+
     cgtcggtgacggctaaggtgaccgggttccacctctgggtcgacccgtacttcgggggtg

       S  H  C  R  P  H  W  P  K  V  E  T  Q  L  G  M  K  P  P  L -

     tcaccagctcagaggccaagaaccacattgccactgatgctgtcagtgctgcagtgggga
3061 ---------+---------+---------+---------+---------+---------+
     agtggtcgagtctccggttcttggtgtaacggtgactacgacagtcacgacgtcacccct

       T  S  S  E  A  K  N  H  I  A  T  D  A  V  S  A  A  V  G  N -

     acctgacaaagccagctctcagtagccccaaggagaaccacggggacttcatcactgatc
3121 ---------+---------+---------+---------+---------+---------+
     tggactgtttcggtcgagagtcatcggggttcctcttggtgcccctgaagtagtgactag

       L  T  K  P  A  L  S  S  P  K  E  N  H  G  D  P  I  T  D  P -

     ccaacgtgtgggtctctgtgcccattgctgaggggaatctgacctcgacgagctcgagg
3181 ---------+---------+---------+---------+---------+---------+
     ggttgcacacccagagacacgggtaacgactccccttagactggagctgctcgagctcc

       N  V  W  V  S  V  P  I  A  E  G  E  S  D  L  D  E  L  E  E -

     aagatatggagcaggcttcgcagagctcctggcaggaagaggaccccaagggacagcagg
3241 ---------+---------+---------+---------+---------+---------+
     ttctatacctcgtccgaagcgtctcgaggaccgtccttctcctggggttccctgtcgtcc

       D  M  E  Q  A  S  Q  S  S  W  Q  E  E  D  P  K  G  Q  Q  E -

     agcagttgccacaagtccaaaagtgtgaaaaccaccaggcagccagaagcccagcctcca
3301 ---------+---------+---------+---------+---------+---------+
     tcgtcaacggtgttcaggttttcacacttttggtggtccgtcggtcttcgggtcggaggt

       Q  L  P  Q  V  Q  K  C  E  N  H  Q  A  A  R  S  P  A  S  M -

     tgatgtcctctgaggacctggctccatacctgggtgagagctggaagaggaaggatagcc
3361 ---------+---------+---------+---------+---------+---------+
     actacaggagactcctggaccgaggtatggacccactctcgaccttctccttcctatcgg

       M  S  S  E  D  L  A  P  Y  L  G  E  S  W  K  R  K  D  S  P -

     ctcaggtccctgccgagggagtggatgacacgagctcctctgagggcagcacggtggact
3421 ---------+---------+---------+---------+---------+---------+
     gagtccagggacggctccctcacctactgtgctcgaggagactcccgtcgtgccacctga

       Q  V  P  A  E  G  V  D  D  T  S  S  S  E  G  S  T  V  D  C -

     gcccggacccagaggaaatcctgaggaagatccccgagctggcacatgacctggacgagc
3481 ---------+---------+---------+---------+---------+---------+
     cgggcctgggtctcctttaggactcctttctaggggctcgaccgtgtactggacctgctcg

       P  D  P  E  E  I  L  R  K  I  P  E  L  A  H  D  L  D  E  P -
```

```
    ccgatgactgtttcagagaaggctgcactcgccgctgtccctgctgcaacgtgaatacta
3541 ---------+---------+---------+---------+---------+---------+
    ggctactgacaaagtctcttccgacgtgagcggcgacagggacgacgttgcacttatgat

     D  D  C  F  R  E  G  C  T  R  R  C  P  C  C  N  V  N  T  S -

    gcaagtctccttgggccacaggctggcaggtgcgcaagacctgctaccgcatcgtggagc
3601 ---------+---------+---------+---------+---------+---------+
    cgttcagaggaacccggtgtccgaccgtccacgcgttctggacgatggcgtagcacctcg

      K  S  P  W  A  T  G  W  Q  V  R  K  T  C  Y  R  I  V  E  H -

    acagctggtttgagagtttcatcatcttcatgatcctgctcagcagtggagcgctggcct

3661 ---------+---------+---------+---------+---------+---------+
    tgtcgaccaaactctcaaagtagtagaagtactaggacgagtcgtcacctcgcgaccgga

      S  W  F  E  S  F  I  I  F  M  I  L  L  S  S  G  A  L  A  F -

    ttgaggataactacctggaagagaaaccccgagtgaagtccgtgctggagtacactgacc
3721 ---------+---------+---------+---------+---------+---------+
    aactcctattgatggaccttctctttggggctcacttcaggcacgacctcatgtgactgg

      E  D  N  Y  L  E  E  K  P  R  V  K  S  V  L  E  Y  T  D  R -

    gagtgttcaccttcatcttcgtctttgagatgctgctcaagtgggtagcctatggcttca
3781 ---------+---------+---------+---------+---------+---------+
    ctcacaagtggaagtagaagcagaaactctacgacgagttcacccatcggataccgaagt

      V  F  T  F  I  F  V  F  E  M  L  L  K  W  V  A  Y  G  F  K -

    aaaagtatttcaccaatgcctggtgctggctggacttcctcattgtgaacatctccctga
3841 ---------+---------+---------+---------+---------+---------+
    ttttcataaagtggttacggaccacgaccgacctgaaggagtaacacttgtagagggact

      K  Y  F  T  N  A  W  C  W  L  D  F  L  I  V  N  I  S  L  T -

    caagcctcatagcgaagatccttgagtattccgacgtggcgtccatcAAAGCCCTTCGGA
3901 ---------+---------+---------+---------+---------+---------+
    gttcggagtatcgcttctaggaactcataaggctgcaccgcaggtagTTTCGGGAAGCCT

      S  L  I  A  K  I  L  E  Y  S  D  V  A  S  I  K  A  L  R  T -

    CTCTCCGTGCCCTCCGACCGCTGCGGGCTCTGTCTCGATTCGAAGGCATGAGGGTAGTGG
3961 ---------+---------+---------+---------+---------+---------+
    GAGAGGCACGGGAGGCTGGCGACGCCCGAGACAGAGCTAAGCTTCCGTACTCCCATCACC

      L  R  A  L  R  P  L  R  A  L  S  R  F  E  G  M  R  V  V  V -

    TGGATGCCCTCGTGGGCGCCATCCCCTCCATCATGAACGTCCTCCTCGTCTGCCTCATCT
4021 ---------+---------+---------+---------+---------+---------+
    ACCTACGGGAGCACCCGCGGTAGGGGAGGTAGTACTTGCAGGAGGAGCAGACGGAGTAGA

      D  A  L  V  G  A  I  P  S  I  M  N  V  L  L  V  C  L  I  F -

    TCTGGCTCATCTTCAGCATCATGGGCGTGAACCTCTTCGCCGGGAAATTTTCGAAGTGCG
4081 ---------+---------+---------+---------+---------+---------+
    AGACCGAGTAGAAGTCGTAGTACCCGCACTTGGAGAAGCGGCCCTTTAAAAGCTTCACGC

      W  L  I  F  S  I  M  G  V  N  L  F  A  G  K  F  S  K  C  V -
```

```
     TCGACACCAGAAATAACCCATTTTCCAACGTGAATTCGACGATGGTGAATAACAAGTCCG
4141 ---------+---------+---------+---------+---------+---------+
     AGCTGTGGTCTTTATTGGGTAAAAGGTTGCACTTAAGCTGCTACCACTTATTGTTCAGGC

       D   T   R   N   N   P   F   S   N   V   N   S   T   M   V   N   N   K   S   E -

     AGTGTCACAATCAAAACAGCACCGGCCACTTCTTCTGGGTCAACGTCAAAGTCAACTTCG
4201 ---------+---------+---------+---------+---------+---------+
     TCACAGTGTTAGTTTTGTCGTGGCCGGTGAAGAAGACCCAGTTGCAGTTTCAGTTGAAGC

       C   H   N   Q   N   S   T   G   H   F   F   W   V   N   V   K   V   N   F   D -

     ACAACGTCGCTATGGGCTACCTCGCACTTCTTCAGGTGGCAACCTTCAAAGGCTGGATGG
4261 ---------+---------+---------+---------+---------+---------+
     TGTTGCAGCGATACCCGATGGAGCGTGAAGAAGTCCACCGTTGGAAGTTTCCGACCTACC

       N   V   A   M   G   Y   L   A   L   L   Q   V   A   T   F   K   G   W   M   D -

     ACATAATGTATGCAGCTGTTGATTCCGGAGAGATCAACAGTCAGCCTAACTGGGAGAACA
4321 ---------+---------+---------+---------+---------+---------+
     TGTATTACATACGTCGACAACTAAGGCCTCTCTAGTTGTCAGTCGGATTGACCCTCTTGT

       I   M   Y   A   A   V   D   S   G   E   I   N   S   Q   P   N   W   E   N   N -

     ACTTGTACATGTACCTGTACTTCGTCGTTTTCATCATTTTCGGTGGCTTCTTCACGCTGA
4381 ---------+---------+---------+---------+---------+---------+
     TGAACATGTACATGGACATGAAGCAGCAAAAGTAGTAAAAGCCACCGAAGAAGTGCGACT

       L   Y   M   Y   L   Y   F   V   V   F   I   I   F   G   G   F   F   T   L   N -

     ATCTCTTTGTTGGGGTCATAATCGACAACTTCAACCAACAGAAAAAAAAGCTAGGAGGCC
4441 ---------+---------+---------+---------+---------+---------+
     TAGAGAAACAACCCCAGTATTAGCTGTTGAAGTTGGTTGTCTTTTTTTTCGATCCTCCGG

       L   F   V   G   V   I   I   D   N   F   N   Q   Q   K   K   K   L   G   G   Q -

     AGGACATCTTCATGACAGAAGAGCAGAAGAAGTACTACAATGCCATGAAGAAGCTGGGCT
4501 ---------+---------+---------+---------+---------+---------+
     TCCTGTAGAAGTACTGTCTTCTCGTCTTCTTCATGATGTTACGGTACTTCTTCGACCCGA

       D   I   F   M   T   E   E   Q   K   K   Y   Y   N   A   M   K   K   L   G   S -

     CCAAGAAACCCCAGAAGCCCATCCCACGGCCCCTGAATAAGTACCAAGGCTTCGTGTTTG
4561 ---------+---------+---------+---------+---------+---------+
     GGTTCTTTGGGGTCTTCGGGTAGGGTGCCGGGGACTTATTCATGGTTCCGAAGCACAAAC

       K   K   P   Q   K   P   I   P   R   P   L   N   K   Y   Q   G   F   V   F   D -

     ACATCGTGACCAGGCAAGCCTTTGACATCATCATCATGGTTCTCATCTGCCTCAACATGA
4621 ---------+---------+---------+---------+---------+---------+
     TGTAGCACTGGTCCGTTCGGAAACTGTAGTAGTAGTACCAAGAGTAGACGGAGTTGTACT

       I   V   T   R   Q   A   F   D   I   I   I   M   V   L   I   C   L   N   M   I -

     TCACCATGATGGTGGAGACCGACGAGCAGGGCGAGGAGAAGACGAAGGTTCTCGGCAGAA
4681 ---------+---------+---------+---------+---------+---------+
     AGTGGTACTACCACCTCTGGCTGCTCGTCCCGCTCCTCTTCTGCTTCCAAGACCCGTCTT

       T   M   M   V   E   T   D   E   Q   G   E   E   K   T   K   V   L   G   R   I -
```

```
     TCAACCAGTTCTTTGTGGCCGTCTTCACGGGCGAGTGTGTGATGAAGATGTTCGCCCTGC
4741 ---------+---------+---------+---------+---------+---------+
     AGTTGGTCAAGAAACACCGGCAGAAGTGCCCGCTCACACACTACTTCTACAAGCGGGACG

       N  Q  F  F  V  A  V  F  T  G  E  C  V  M  K  M  F  A  L  R -

     GACAGTACTACTTCACCAACGGCTGGAACGTGTTCGACTTCATAGTGGTGATCCTGTTCA
4801 ---------+---------+---------+---------+---------+---------+
     CTGTCATGATGAAGTGGTTGCCGACCTTGCACAAGCTGAAGTATCACCACTAGGACAGGT

       Q  Y  Y  F  T  N  G  W  N  V  F  D  F  I  V  V  I  L  S  I -

     TTGGGAGTCTGCTGTTTTCTGCAATCCTTAAGTCACTGGAAAACTACTTCTCCCCGACGC
4861 ---------+---------+---------+---------+---------+---------+
     AACCCTCAGACGACAAAAGACGTTAGGAATTCAGTGACCTTTTGATGAAGAGGGGCTGCG

       G  S  L  L  F  S  A  I  L  K  S  L  E  N  Y  F  S  P  T  L -

     TCTTCCGGGTCATCCGTCTGGCCAGGATCGGCCGCATCCTCAGGCTGATCCGAGCAGCCA
4921 ---------+---------+---------+---------+---------+---------+
     AGAAGGCCCAGTAGGCAGACCGGTCCTAGCCGGCGTAGGAGTCCGACTAGGCTCGTCGGT

       F  R  V  I  R  L  A  R  I  G  R  I  L  R  L  I  R  A  A  K -

     AGGGGATTCGCACGCTGCTCTTCGCCCTCATGATGTCCCTGCCCGCCCTCTTCAACATCG
4981 ---------+---------+---------+---------+---------+---------+
     TCCCCTAAGCGTGCGACGAGAAGCGGGAGTACTACAGGGACGGGCGGGAGAAGTTGTAGC

       G  I  R  T  L  L  F  A  L  M  M  S  L  P  A  L  F  N  I  G -

     GCCTCCTCCTCTTCCTCGTCATGTTCATCTACTCCATCTTCGGCATGGCCAGCTTCGCTA
5041 ---------+---------+---------+---------+---------+---------+
     CGGAGGAGGAGAAGGAGCAGTACAAGTAGATGAGGTAGAAGCCGTACCGGTCGAAGCGAT

       L  L  L  F  L  V  M  F  I  Y  S  I  F  G  M  A  S  F  A  N -

     ACGTCGTGGACGAGGCCGGCATCGACGACATGTTCAACTTCAAGACCTTTGGCAACAGCA
5101 ---------+---------+---------+---------+---------+---------+
     TGCAGCACCTGCTCCGGCCGTAGCTGCTGTACAAGTTGAAGTTCTGGAAACCGTTGTCGT

       V  V  D  E  A  G  I  D  D  M  F  N  F  K  T  F  G  N  S  M -

     TGCTGTGCCTGTTCCAGATCACCACCTCGGCCGGCTGGGACGGCCTCCTCAGCCCCATCC
5161 ---------+---------+---------+---------+---------+---------+
     ACGACACGGACAAGGTCTAGTGGTGGAGCCGGCCGACCCTGCCGGAGGAGTCGGGGTAGG

       L  C  L  F  Q  I  T  T  S  A  G  W  D  G  L  L  S  P  I  L -

     TCAACACGGGGCCTCCCTACTGCGACCCCAACCTGCCCAACAGCAACGGCTCCCGGGGGA
5221 ---------+---------+---------+---------+---------+---------+
     AGTTGTGCCCCGGAGGGATGACGCTGGGGTTGGACGGGTTGTCGTTGCCGAGGGCCCCCT

       N  T  G  P  P  Y  C  D  P  N  L  P  N  S  N  G  S  R  G  N -

     ACTGCGGGAGCCCCGGCGGTGGGCATCATCTTCTTCACCACCTACATCATCATCTCCTTCC
5281 ---------+---------+---------+---------+---------+---------+
     TGACGCCCTCGGGCCGCCACCCGTAGTAGAAGAAGTGGTGGATGTAGTAGTAGAGGAAGG

       C  G  S  P  A  V  G  I  I  F  F  T  T  Y  I  I  I  S  F  L -
```

```
     TCATCGTGGTCAACATGTACATCGCAGTGATTCTGGAGAACTTCAACGTAGCCACCGAGG
5341 ---------+---------+---------+---------+---------+---------+
     AGTAGCACCAGTTGTACATGTACCGTCACTAAGACCTCTTGAAGTTGCATCGGTGGCTCC

      I  V  V  N  M  Y  I  A  V  I  L  E  N  F  N  V  A  T  E  E -

     AGAGCACGGAGCCCCTGAGCGAGGACGACTTCGACATGTTCTATGAGACCTGGGAGAAGT
5401 ---------+---------+---------+---------+---------+---------+
     TCTCGTGCCTCGGGGACTCGCTCCTGCTGAAGCTGTACAAGATACTCTGGACCCTCTTCA

      S  T  E  P  L  S  E  D  D  F  D  M  F  Y  E  T  W  E  K  F -

     TCGACCCGGAGGCCACCCAGTTCATTGCCTTTTCTGCCCTCTCAGACTTCGCGGACACGC
5461 ---------+---------+---------+---------+---------+---------+
     AGCTGGGCCTCCGGTGGGTCAAGTAACGGAAAAGACGGGAGAGTCTGAAGCGCCTGTGCG

      D  P  E  A  T  Q  F  I  A  F  S  A  L  S  D  F  A  D  T  L -

     TCTCCGGCCCTCTTAGAATCCCCAAACCCAACCAGAATATATTAATCCAGATGGACCTGC
5521 ---------+---------+---------+---------+---------+---------+
     AGAGGCCGGGAGAATCTTAGGGGTTTGGGTTGGTCTTATATAATTAGGTCTACCTGGACG

      S  G  P  L  R  I  P  K  P  N  Q  N  I  L  I  Q  M  D  L  P -

     CGTTGGTCCCCGGGGATAAGATCCACTGTCTGGACATCCTTTTTGCCTTCACAAAGAACG
5581 ---------+---------+---------+---------+---------+---------+
     GCAACCAGGGGCCCCTATTCTAGGTGACAGACCTGTAGGAAAAACGGAAGTGTTTCTTGC

      L  V  P  G  D  K  I  H  C  L  D  I  L  F  A  F  T  K  N  V -

     TCTTGGGAGAATCCGGGGAGTTGGACTCCCTGAAGACCAATATGGAAGAGAAGTTTATGG
5641 ---------+---------+---------+---------+---------+---------+
     AGAACCCTCTTAGGCCCCTCAACCTGAGGGACTTCTGGTTATACCTTCTCTTCAAATACC

      L  G  E  S  G  E  L  D  S  L  K  T  N  M  E  E  K  F  M  A -

     CGACCAATCTCTCCAAAGCATCCTATGAACCAATAGCCACCACCCTCCGGTGGAAGCAGG
5701 ---------+---------+---------+---------+---------+---------+
     GCTGGTTAGAGAGGTTTCGTAGGATACTTGGTTATCGGTGGTGGGAGGCCACCTTCGTCC

      T  N  L  S  K  A  S  Y  E  P  I  A  T  T  L  R  W  K  Q  E -

     AAGACCTCTCAGCCACAGTCATTCAAAAGGCCTACCGGAGCTACATGCTGCACCGCTCCT
5761 ---------+---------+---------+---------+---------+---------+
     TTCTGGAGAGTCGGTGTCAGTAAGTTTTCCGGATGGCCTCGATGTACGACGTGGCGAGGA

      D  L  S  A  T  V  I  Q  K  A  Y  R  S  Y  M  L  E  R  S  L -

     TGACACTCTCCAACACCCTGCATGTGCCCAGGGCTGAGGAGGATGGCGTGTCACTTCCCG
5821 ---------+---------+---------+---------+---------+---------+
     ACTGTGAGAGGTTGTGGGACGTACACGGGTCCCGACTCCTCCTACCGCACAGTGAAGGGC

      T  L  S  N  T  L  H  V  P  R  A  E  E  D  G  V  S  L  P  G -

     GGGAAGGCTACATTACATTCATGGCAAACAGTGGACTCCCGGACAAATCAGAAACTGCCT
5881 ---------+---------+---------+---------+---------+---------+
     CCCTTCCGATGTAATGTAAGTACCGTTTGTCACCTGAGGGCCTGTTTAGTCTTTGACGGA

      E  G  Y  I  T  F  M  A  N  S  G  L  P  D  K  S  E  T  A  S -

     CTGCTACGTCTTTCCCGCCATCCTATGACAGTGTCACCAGGGGCCTGAGTGACCGGGCCA
```

```
5941 ---------+---------+---------+---------+---------+---------+
     GACGATGCAGAAAGGCCGGTAGGATACTGTCACAGTGGTCCCCGGACTCACTGGCCCGGT

       A   T   S   F   P   P   S   Y   D   S   V   T   R   G   L   S   D   R   A   N -

     ACATTAACCCATCTAGCTCAATGCAAAATGAAGATGAGGTCGCTGCTAAGGAAGGAAACA
6001 ---------+---------+---------+---------+---------+---------+
     TGTAATTGGGTAGATCGAGTTACGTTTTACTTCTACTCCAGCGACGATTCCTTCCTTTGT

       I   N   P   S   S   S   M   Q   N   E   D   E   V   A   A   K   E   G   N   S -

     GCCCTGGACCTCAGTGAAGGCACTCAGGCATGCACAGGGCAGGTTCCAATGTCTTTCTCT
6061 ---------+---------+---------+---------+---------+---------+
     CGGGACCTGGAGTCACTTCCGTGAGTCCGTACGTGTCCCGTCCAAGGTTACAGAAAGAGA

       P   G   P   Q   *   R   H   S   G   M   H   R   A   G   S   N   V   F   L   C -

     GCTGTACTAACTCCTTCCCTCTGGAGGTGGCACCAACCTCCAGCCTCCACCAATGCATGT
6121 ---------+---------+---------+---------+---------+---------+
     CGACATGATTGAGGAAGGGAGACCTCCACCGTGGTTGGAGGTCGGAGGTGGTTACGTACA

       C   T   N   S   F   P   L   E   V   A   P   T   S   S   L   H   Q   C   M   S -

     CACTGGTCATGGTGTCAGAACTGAATGGGGACATCCTTGAGAAAGCCCCCACCCCAATAG
6181 ---------+---------+---------+---------+---------+---------+
     GTGACCAGTACCACAGTCTTGACTTACCCCTGTAGGAACTCTTTCGGGGGGTGGGGTTATC

       L   V   M   V   S   E   L   N   G   D   I   L   E   K   A   P   T   P   I   G -

     GAATCAAAAGCCAAGGATACTCCTCCATTCTGACGTCCCTTCCGAGTTCCCAGAAGATGT
6241 ---------+---------+---------+---------+---------+---------+
     CTTAGTTTTCGGTTCCTATGAGGAGGTAAGACTGCAGGGAAGGCTCAAGGGTCTTCTACA

       I   K   S   Q   G   Y   S   S   I   L   T   S   L   P   S   S   Q   K   M   S -

     CATTGCTCCCTTCTGTTTGTGACCAGAGACGTGATTCACCAACTTCTCGGAGCCAGAGAC
6301 ---------+---------+---------+---------+---------+---------+
     GTAACGAGGGAAGACAAACACTGGTCTCTGCACTAAGTGGTTGAAGAGCCTCGGTCTCTG

       L   L   P   S   V   C   D   Q   R   R   D   S   P   T   S   R   S   Q   R   H -

     ACATAGCAAAGACTTTTCTGCTGGTGTCGGGCAGTCTTAGAGAAGTCACGTAGGGGTTGG
6361 ---------+---------+---------+---------+---------+---------+
     TGTATCGTTTCTGAAAAGACGACCACAGCCCGTCAGAATCTCTTCAGTGCATCCCCAACC

       I   A   K   T   F   L   L   V   S   G   S   L   R   E   V   T

     TACTGAGAATTAGGGTTTGCATGACTGCATGCTCACAGCTGCCGGACAATACCTGTGAGT
6421 ---------+---------+---------+---------+---------+---------+
     ATGACTCTTAATCCCAAACGTACTGACGTACGAGTGTCGACGGCCTGTTATGGACACTCA



     CGGCCATTAAAATTAATATTTTTAAAGTTAAAAAAAAAAAAAAAA
6481 ---------+---------+---------+---------+---- 6524
     GCCGGTAATTTTAATTATAAAAATTTCAATTTTTTTTTTTTTTT
```

Figure 1b

Nucleic acid and amino acid sequence of variant TTXi DRG sodium channel SNS-C (numbering is the same as SNS-B)

```
ggaaacggtacccacgaaggcgtgattcctcgtagctccctgagtgaggttccctcaggca
----------+---------+---------+---------+---------+---------+
cctttgccatgggtgcttccgcactaaggagcatcgagggactcactccaagggagtccgt

tcaggcaagccccggccaccttcgtgggaagcgcggacaagt
----------+---------+---------+---------+
agtccgttcgggccggtggaagcaccctteqcgcctgttca
```

```
                                                    ATGGAGCTCCCCTTTGCGTCCGTGGGAACTACCAATT
                                                    ------+---------+---------+---------+
                                                    TACCTCGAGGGCAAACGCAGGCACCCTTGATGGTTAA

                                                    M  E  L  P  F  A  S  V  G  T  T  N  F -

    TCAGACGGTTCACTCCAGAGTCACTGGCAGAGATCGAGAAGCAGATTGCTGCTCACCGCG
241 ---------+---------+---------+---------+---------+---------+
    AGTCTGCCAAGTGAGGTCTCAGTGACCGTCTCTAGCTCTTCGTCTAACGACGAGTGGCGC

      R  R  F  T  P  E  S  L  A  E  I  E  K  Q  I  A  A  H  R  A -

    CAGCCAAGAAGGCCAGAACCAAGCACAGAGGACAGGAGGACAAGGGCGAGAAGCCCAGGC
301 ---------+---------+---------+---------+---------+---------+
    GTCGGTTCTTCCGGTCTTGGTTCGTGTCTCCTGTCCTCCTGTTCCCGCTCTTCGGGTCCG

     A  K  K  A  R  T  K  H  R  G  Q  E  D  K  G  E  K  P  R  P -

    CTCAGCTGGACTTGAAAGACTGTAACCAGCTGCCCAAGTTCTATGGTGAGCTCCCAGCAG
361 ---------+---------+---------+---------+---------+---------+
    GAGTCGACCTCAACTTTCTGACATTGGTCGACGGGTTCAAGATACCACTCGAGGGTCGTC

     Q  L  D  L  K  D  C  N  Q  L  P  K  F  Y  G  E  L  P  A  E -

    AACTGGTCGGGGAGCCCCTGGAGGACCTAGACCCTTTCTACAGCACACACCGGACATTCA
421 ---------+---------+---------+---------+---------+---------+
    TTGACCAGCCCCTCGGGGACCTCCTGGATCTGGGAAAGATGTCGTGTGTGGCCTGTAAGT

      L  V  G  E  P  L  E  D  L  D  P  F  Y  S  T  H  R  T  F  M -

    TGGTGTTGAATAAAAGCAGGACCATTTCCAGATTCAGTGCCACTTGGGCCCTGTGGCTCT
481 ---------+---------+---------+---------+---------+---------+
    ACCACAACTTATTTTCGTCCTGGTAAAGGTCTAAGTCACGGTGAACCCGGGACACCGAGA

     V  L  N  K  S  R  T  I  S  R  F  S  A  T  W  A  L  W  L  P -

    TCAGTCCCTTCAACCTGATCAGAAGAACAGCCATCAAAGTGTCTGTCCATTCCTGGTTCT
541 ---------+---------+---------+---------+---------+---------+
    AGTCAGGGAAGTTGGACTAGTCTTCTTGTCGGTAGTTTCACAGACAGGTAAGGACCAAGA
```

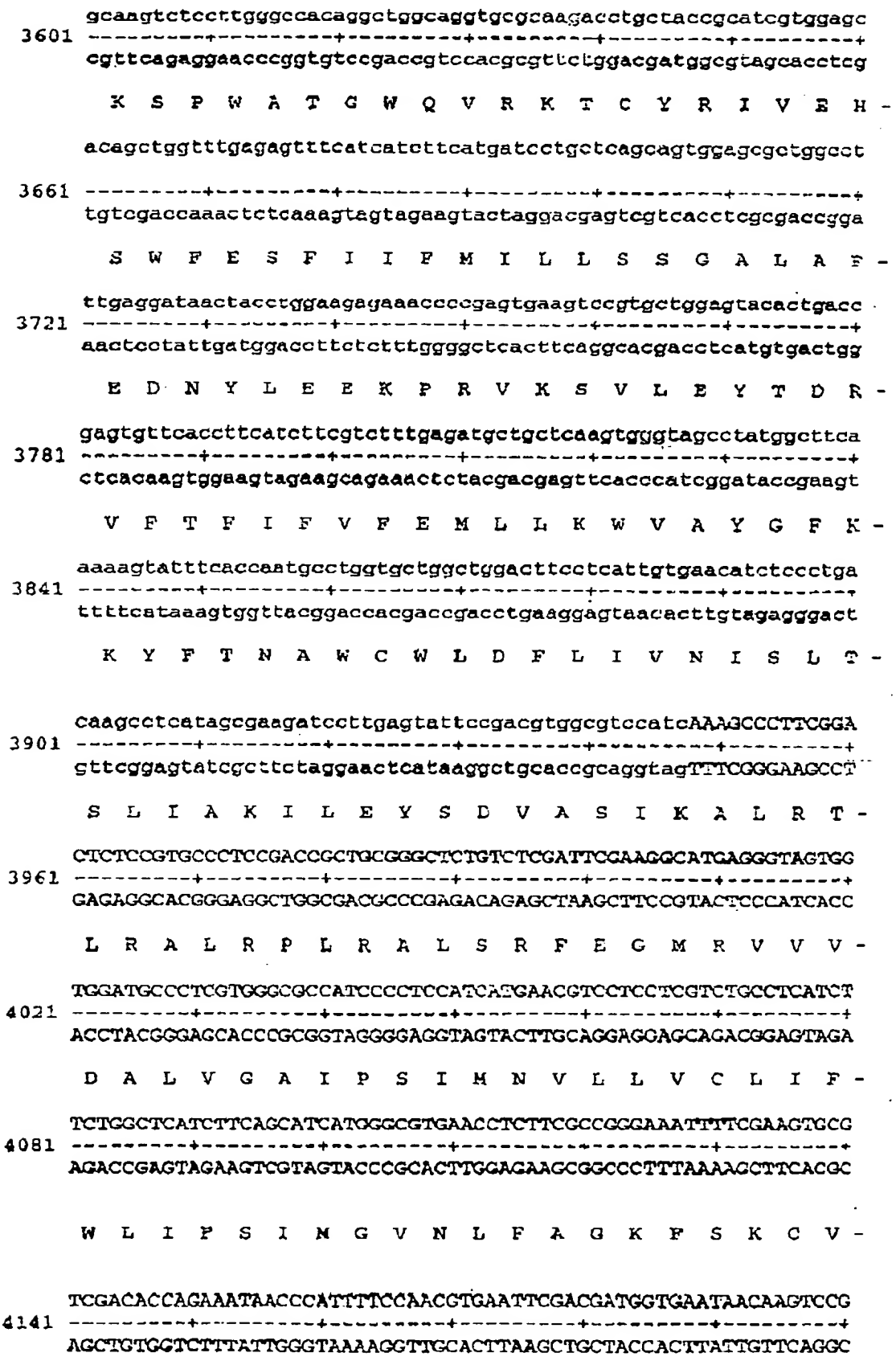
```
      S  P  F  N  L  I  R :R  T  A  I  K  V  S  V  H  S  W  F  S -

     CCATATTCATCACCATCACTATTTTGGTCAACTGCGTGTGCATGACCCGAACTGATCTTC
 601 ---------+---------+---------+---------+---------+---------+
     GGTATAAGTAGTGGTAGTGATAAAACCAGTTGACGCACACGTACTGGGCTTGACTAGAAG

      I  F  I  T  I  T  I  L  V  N  C  V  C  M  T  R  T  D  L  P -

     CAGAGAAAGTCGAGTACGTCTTCACTGTCATTTACACCTTCGAGGCTCTGATTAAGATAC
 661 ---------+---------+---------+---------+---------+---------+
     GTCTCTTTCAGCTCATGCAGAAGTGACAGTAAATGTGGAAGCTCCGAGACTAATTCTATG

      E  K  V  E  Y  V  F  T  V  I  Y  T  F  E  A  L  I  K  I  L -

     TGGCAAGAGGGTTTTGTCTAAATGAGTTCACTTATCTTCGAGATCCGTGGAACTGGCTGG
 721 ---------+---------+---------+---------+---------+---------+
     ACCGTTCTCCCAAAACAGATTTACTCAAGTGAATAGAAGCTCTAGGCACCTTGACCGACC

      A  R  G  F  C  L  N  E  F  T  Y  L  R  D  P  W  N  W  L  D -

     ACTTCAGTGTCATTACCTTGGCGTATGTGGGTGCAGCGATAGACCTCCGAGGAATCTCAG
 781 ---------+---------+---------+---------+---------+---------+
     TGAAGTCACAGTAATGGAACCGCATACACCCACGTCGCTATCTGGAGGCTCCTTAGAGTC

      F  S  V  I  T  L  A  Y  V  G  A  A  I  D  L  R  G  I  S  G -

     GCCTGCGGACATTCCGAGttctcagagccctgaaaactgtttctgtgatcccaggactga
 841 ---------+---------+---------+---------+---------+---------+
     CGGACGCCTGTAAGGCTCaagagtctcgggacttttgacaaagacactagggtcctgact

      L  R  T  F  R  V  L  R  A  L  K  T  V  S  V  I  P  G  L  K -

     aggtcatcgtgggagccctgatccactcagtgaggaagctggccgacgtgactatcctca
 901 ---------+---------+---------+---------+---------+---------+
     tccagtagcaccctcgggactaggtgagtcactccttcgaccggctgcactgataggagt

      V  I  V  G  A  L  I  H  S  V  R  K  L  A  D  V  T  I  L  T -

     cagtcttctgcctgagcgtcttcgccttggtgggcctgcagctctttaaggggaacctta
 961 ---------+---------+---------+---------+---------+---------+
     gtcagaagacggactcgcagaagcggaaccacccggacgtcgagaaattccccttggaat

      V  F  C  L  S  V  F  A  L  V  G  L  Q  L  F  K  G  N  L  K -

     agaacaaatgcatcaggaacggaacagatccccacaaggctgacaacctctcatctgaaa
1021 ---------+---------+---------+---------+---------+---------+
     tcttgtttacgtagtccttgccttgtctaggggtgttccgactgttggagagtagacttt

      N  K  C  I  R  N  G  T  D  P  H  K  A  D  N  L  S  S  E  M -

     tggcagaatacgtctccatcaagcctggtactacggatcccttactgtgcggcaatgggt
1081 ---------+---------+---------+---------+---------+---------+
     accgtcttatgcagaggtagttcggaccatgatgcctagggaatgacacgccgttaccca

      A  E  Y  V  S  I  K  P  G  T  T  D  P  L  L  C  G  N  G  S -

     ctgatgctggtcactgccctggaggctatgtctgcctgaaaactcctgacaacccggatt
1141 ---------+---------+---------+---------+---------+---------+
     gactacgaccagtgacgggacctccgatacagacggacttttgaggactgttgggcctaa
```

```
      D  A  G  H  C  P  G  G  Y  V  C  L  K  T  P  D  N  P  D  F -

      ttaactacaccagctttgattcctttgcgtgggcattcctctcactgttccgcctcatga
1201  ---------+---------+---------+---------+---------+---------+
      aattgatgtggtcgaaactaaggaaacgcacccgtaaggagagtgacaaggcggagtact

      N  Y  T  S  F  D  S  F  A  W  A  F  L  S  L  F  R  L  M  T -

      cgcaggactcctgggagcgcctgtaccagcagacactccgggcttctgggaaaatgtaca
1261  ---------+---------+---------+---------+---------+---------+
      gcgtcctgaggaccctcgcggacatggtcgtctgtgaggcccgaagacccttttacatgt

       Q  D  S  W  E  R  L  Y  Q  Q  T  L  R  A  S  G  K  M  Y  M -

      tggtcttttcgtgctggttattttccttggatcgttctacctggtcaatttgatcttgg
1321  ---------+---------+---------+---------+---------+---------+
      accagaaaaagcacgaccaataaaaggaacctagcaagatggaccagttaaactagaacc

       V  P  F  V  L  V  I  F  L  G  S  F  Y  L  V  N  L  I  L  A -

      ccgtggtcaccatggcgtatgaagagcagagccaggcaacaattgcagaaatcgaagcca
1381  ---------+---------+---------+---------+---------+---------+
      ggcaccagtggtaccgcatacttctcgtctcggtccgttgttaacgtctttagcttcggt

       V  V  T  M  A  Y  E  E  Q  S  Q  A  T  I  A  E  I  E  A  K -

      aggaaaaaaagttccaggaagcccttgaggtgctgcagaaggaacaggaggtgctggcag
1441  ---------+---------+---------+---------+---------+---------+
      tcctttttttcaaggtccttcgggaactccacgacgtcttccttgtcctccacgaccgtc

       E  K  K  F  Q  E  A  L  E  V  L  Q  K  E  Q  E  V  L  A  A -

      cctggggattgacacgacctcgctccagtcccacagtggatcacccttagcctccaaaa
1501  ---------+---------+---------+---------+---------+---------+
      gggaccccctaactgtgctggagcgaggtcagggtgtcacctagtgggaatcggaggtttt

       L  G  I  D  T  T  S  L  Q  S  H  S  G  S  P  L  A  S  K  N -

      acgccaatgagagaagacccagggtgaaatcaagggtgtcagagggctccacggatgaca
1561  ---------+---------+---------+---------+---------+---------+
      tgcggttactctcttctgggtcccactttagttcccacagtctcccgaggtgcctactgt

       A  N  E  R  R  P  R  V  K  S  R  V  S  E  G  S  T  D  D  N -

      acaggtcaccccaatctgacccttacaaccagcgcaggatgtctttcctaggcctgtctt
1621  ---------+---------+---------+---------+---------+---------+
      tgtccagtggggttagactgggaatgttggtcgcgtcctacagaaaggatccggacagaa

       R  S  P  Q  S  D  P  Y  N  Q  R  R  M  S  F  L  G  L  S  S -

      caggaagacgcagggctagccacggcagtgtgttccacttccgagcgcccagccaagaca
1681  ---------+---------+---------+---------+---------+---------+
      gtccttctgcgtcccgatcggtgccgtcacacaaggtgaaggctcgcgggtcggttctgt

       G  R  R  R  A  S  H  G  S  V  F  H  F  R  A  P  S  Q  D  I -

      tctcatttcctgacgggatcacccctgatgatggggtctttcacggagaccaggaaagcc
1741  ---------+---------+---------+---------+---------+---------+
      agagtaaaggactgccctagtggggactactaccccagaaagtgcctctggtcctttcgg
```

```
      S  F  P  D  G  I  T  P  D  D  G  V  F  H  G  D  Q  E  S  R -

      gtcgaggttccatattgctgggcaggggtgctgggcagacaggtccactccccaggagcc
1801  ---------+---------+---------+---------+---------+---------+
      cagctccaaggtataacgacccgtccccacgacccgtctgtccaggtgaggggtcctcgg

      R  G  S  I  L  L  G  R  G  A  C  Q  T  G  P  L  P  R  S  P -

      cactgcctcagtcccccaaccctggccgtagacatggagaaggagggacagctcggagtgc
1861  ---------+---------+---------+---------+---------+---------+
      gtgacggagtcaggggggttgggaccggcatctgtacctcttctcctgtcgagcctcacg

       L  P  Q  S  P  N  P  G  R  R  H  G  E  E  G  Q  L  G  V  P -

      ccactggtgagcttaccgctggagcgcctgaaggcccggcactgcacactacagggcaga
1921  ---------+---------+---------+---------+---------+---------+
      ggtgaccactcgaatggcgacctcgcggacttccgggccgtgacgtgtgatgtcccgtct

      T  G  E  L  T  A  G  A  P  E  G  P  A  L  H  T  T  G  Q  K -

      agagcttcctgtctgcgggctacttgaacgaacctttccgagcacagagggccatgagcg
1981  ---------+---------+---------+---------+---------+---------+
      tctcgaaggacagacgcccgatgaacttgcttggaaaggctcgtgtctcccggtactcgc

      S  F  L  S  A  G  Y  L  N  E  P  F  R  A  Q  R  A  M  S  V -

      ttgtcagtatcatgacttctgtcattgaggcgcttgaagagtctaagctgaagtgcccac
2041  ---------+---------+---------+---------+---------+---------+
      aacagtcatagtactgaagacagtaactcctcgaacttctcagattcgacttcacgggtg

       V  S  I  M  T  S  V  I  E  E  L  E  E  S  K  L  K  C  P  P -

      cctgcttgatcagcttcgctcagaagtatctgatctgggagtgctgccccaagtggaggga
2101  ---------+---------+---------+---------+---------+---------+
      ggacgaactagtcgaagcgagtcttcatagactagaccctcacgacggggttcacctcct

      C  L  I  S  P  A  Q  K  Y  L  I  W  E  C  C  P  K  W  R  K -

      agttcaagatggcgctgttcgagctggtgactgacccttcgcagagcttaccatcaccc
2161  ---------+---------+---------+---------+---------+---------+
      tcaagttctaccgcgacaagctcgaccactgactggggaagcgtctcgaatggtagtggg

      F  K  M  A  L  F  E  L  V  T  D  P  F  A  E  L  T  I  T  L -

      tctgcatcgtggtgaacaccgtcttcatggccatggagcactaccccatgaccgatgcct
2221  ---------+---------+---------+---------+---------+---------+
      agacgtagcaccacttgtggcagaagtaccggtacctcgtgatggggtactggctacgga

      C  I  V  V  N  T  V  F  M  A  M  E  H  Y  P  M  T  D  A  P -

      tcgatgccatgcttcaagccggcaacattgtcttcaccgtgtttttcacaatggagatgg
2281  ---------+---------+---------+---------+---------+---------+
      agctacggtacgaagttcggccgttgtaacagaagtggcacaaaaagtgttaccctctacc

      D  A  M  L  Q  A  G  N  I  V  F  T  V  F  F  T  M  E  M  A -

      ccttcaagatcattgccttcgacccctactattacttccagaagaagtggaatatcttcg
2341  ---------+---------+---------+---------+---------+---------+
      ggaagttctagtaacggaagctggggatgataatgaaggtcttcttcaccttatagaagc
```

```
      F  K  I  I  A  F  D  P  Y  Y  Y  F  Q  K  K  W  N  I  F  D -

      actgtgtcatcgtcaccgtgagccttctggagctgagtgcatccaagaagggcagcctgt
2401  ---------+---------+---------+---------+---------+---------+
      tgacacagtagcagtggcactcggaagacctcgactcacgtaggttcttcccgtcggaca

      C  V  I  V  T  V  S  L  L  E  L  S  A  S  K  R  G  S  L  S -

      ctgtgctccgtaccttacgcttgctgcgggtcttcaagctggccaagtcctggcccaccc
2461  ---------+---------+---------+---------+---------+---------+
      gacacgaggcatggaatgcgaacgacgcccagaagttcgaccggttcaggaccgggtggg

      V  L  R  T  L  R  L  L  R  V  F  K  L  A  K  S  W  P  T  L -

      tgaacaccctcatcaagatcatcgggaactcagtgggggccctgggcaacctgaccttta
2521  ---------+---------+---------+---------+---------+---------+
      acttgtgggagtagttctagtagcccttgagtcacccccgggacccgttggactggaaat

      N  T  L  I  K  I  I  G  N  S  V  G  A  L  G  N  L  T  F  Y -

      tcctggccatcatcgtcttcatcttcgccctggtcggaaagcagcttctctcagaggact
2581  ---------+---------+---------+---------+---------+---------+
      aggaccggtagtagcagaagtagaagcgggaccagcctttcgtcgaagagagtctcctga

      L  A  I  I  V  F  I  F  A  L  V  G  K  Q  L  L  S  E  D  Y -

      acgggtgccgcaaggacggcgtctccgtgtggaacggcgagaagctccgctggcacatgt
2641  ---------+---------+---------+---------+---------+---------+
      tgcccacggcgttcctgccgcagaggcacaccttgccgctcttcgaggcgaccgtgtaca

      G  C  R  K  D  G  V  S  V  W  N  G  E  K  L  R  W  H  M  C -

      gtgacttcttccattccttcctggtcgtcttccgaatcctctgcggggagtggatcgaga
2701  ---------+---------+---------+---------+---------+---------+
      cactgaagaaggtaaggaaggaccagcagaaggcttaggagacgcccctcacctagctct

      D  F  F  H  S  F  L  V  V  F  R  I  L  C  G  E  W  I  E  N -

      acatgtgggtctgcatggaggtcagccagaaatccatctgcctcatcctcttcttgactg
2761  ---------+---------+---------+---------+---------+---------+
      tgtacacccagacgtacctccagtcggtctttaggtagacggagtaggagaagaactgac

      M  W  V  C  M  E  V  S  Q  K  S  I  C  L  I  L  F  L  T  V -

      tgatggtgctgggcaacctagtggtgctcaacctttcatcgctttactgctgaactcct
2821  ---------+---------+---------+---------+---------+---------+
      actaccacgacccgttggatcaccacgagttggaaaagtagcgaaatgacgacttgagga

      M  V  L  G  N  L  V  V  L  N  L  F  I  A  L  L  L  N  S  F -

      tcagcgcggacaacctcacggctccagaggatgacggggaggtgaacaacttgcagttag
2881  ---------+---------+---------+---------+---------+---------+
      agtcgcgcctgttggagtgccgaggtctcctactgccccctccacttgttgaacgtcaatc

      S  A  D  N  L  T  A  P  E  D  D  G  E  V  N  N  L  Q  L  A -

      cactggccaggatccaggtacttggccatcgggccagcagggccagcgccagttacatca
2941  ---------+---------+---------+---------+---------+---------+
      gtgaccggtcctaggtccatgaaccggtagcccggtcgtcccggtcgcggtcaatgtagt

      L  A  R  I  Q  V  L  G  H  R  A  S  R  A  S  A  S  Y  I  S -
```

```
      gcagccactgccgattccactggcccaaggtggagacccagctgggcatgaagccccac

3001  ---------+---------+---------+---------+---------+---------+
      cgtcggtgacggctaaggtgaccgggttccacctctgggtcgacccgtacttcgggggtg

       S  H  C  R  F  H  W  P  K  V  E  T  Q  L  G  M  K  P  P  L -

      tcaccagctcagaggccaagaaccacattgccactgatgctgtcagtgctgcagtgggga
3061  ---------+---------+---------+---------+---------+---------+
      agtggtcgagtctccggttcttggtgtaacggtgactacgacagtcacgacgtcaccect

        T  S  S  E  A  K  N  H  I  A  T  D  A  V  S  A  A  V  G  N -

      acctgacaaagccagctctcagtagccccaaggagaaccacggggacttcatcactgatc
3121  ---------+---------+---------+---------+---------+---------+
      tggactgtttcggtcgagagtcatcggggttcctcttggtgccccctgaagtagtgactag

        L  T  K  P  A  L  S  S  P  K  E  N  H  G  D  F  I  T  D  P -

      ccaacgtgtgggtctctgtgcccattgctgaggggaatctgacctcgacgagctcgagg
3181  ---------+---------+---------+---------+---------+---------+
      ggttgcacacccagagacacgggtaacgactccccttagactggagctgctcgagctcc

        N  V  W  V  S  V  P  I  A  E  G  E  S  D  L  D  E  L  E  E -


      aagatatggagcaggcttcgcagagctcctggcaggaagaggaccccaagggacagcagg
3241  ---------+---------+---------+---------+---------+---------+
      ttctatacctcgtccgaagcgtctcgaggaccgtccttctcctggggttccctgtcgtcc

        D  M  E  Q  A  S  Q  S  S  W  Q  E  E  D  P  K  G  Q  Q  E -

      agcagttgccacaagtccaaaagtgtgaaaaccaccaggcagccagaagcccagcctcca
3301  ---------+---------+---------+---------+---------+---------+
      tcgtcaacggtgttcaggttttcacacttttggtggtccgtcggtcttcgggtcggaggt

        Q  L  P  Q  V  Q  K  C  E  N  K  Q  A  A  R  S  P  A  S  M -

      tgatgtcctctgaggacctggctccatacctgggtgagagctggaagaggaaggatagcc
3361  ---------+---------+---------+---------+---------+---------+
      actacaggagactcctggaccgaggtatggacccactctcgaccttctccttcctatcgg

        M  S  S  E  D  L  A  P  Y  L  G  E  S  W  K  R  K  D  S  P -

      ctcaggtccctgccgagggagtggatgacacgagctcctctgagggcagcacggtggact
3421  ---------+---------+---------+---------+---------+---------+
      gagtccagggacggctccctcacctactgtgctcgaggagactcccgtcgtgccacctga

        Q  V  P  A  E  G  V  D  D  T  S  S  S  E  G  S  T  V  D  C -

      gcccggacccagaggaaatcctgaggaagatccccgagctggcacatgacctggacgagc
3481  ---------+---------+---------+---------+---------+---------+
      cgggcctgggtctcctttaggactccttctaggggctcgaccgtgtactggacctgctcg

        P  D  P  E  E  I  L  R  K  I  P  E  L  A  H  D  L  D  E  P -

      ccgatgactgtttcagagaaggctgcactcgccgctgtccctgctgcaacgtgaatacta
3541  ---------+---------+---------+---------+---------+---------+
      ggctactgacaaagtctcttccgacgtgagcggcgacagggacgacgttgcacttatgat

        D  D  C  F  R  E  G  C  T  R  R  C  P  C  C  N  V  N  T  S -
```


gcaagtctccttgggccacaggctggcaggtgcgcaagacctgctaccgcatcgtggagc
3601
cgttcagaggaacccggtgtccgaccgtccacgcgttctggacgatggcgtagcacctcg
K S P W A T G W Q V R K T C Y R I V E H -
acagctggtttgagagtttcatcatcttcatgatcctgctcagcagtggagcgctggcct
3661
tgtcgaccaaactctcaaagtagtagaagtactaggacgagtcgtcacctcgcgaccgga
S W F E S F I I F M I L L S S G A L A F -
ttgaggataactacctggaagagaaaccccgagtgaagtccgtgctggagtacactgacc
3721
aactcctattgatggaccttctctttggggctcacttcaggcacgacctcatgtgactgg
E D N Y L E E K P R V K S V L E Y T D R -
gagtgttcaccttcatcttcgtctttgagatgctgctcaagtgggtagcctatggcttca
3781
ctcacaagtggaagtagaagcagaaactctacgacgagttcacccatcggataccgaagt
V F T F I F V F E M L L K W V A Y G F K -
aaaagtatttcaccaatgcctggtgctggctggacttcctcattgtgaacatctccctga
3841
ttttcataaagtggttacggaccacgaccgacctgaaggagtaacacttgtagagggact
K Y F T N A W C W L D F L I V N I S L T -
caagcctcatagcgaagatccttgagtattccgacgtggcgtccatcAAAGCCCTTCGGA
3901
gttcggagtatcgcttctaggaactcataaggctgcaccgcaggtagTTTCGGGAAGCCT
S L I A K I L E Y S D V A S I K A L R T -
CTCTCCGTGCCCTCCGACCGCTGCGGGCTCTGTCTCGATTCGAAGGCATGAGGGTAGTGG
3961
GAGAGGCACGGGAGGCTGGCGACGCCCGAGACAGAGCTAAGCTTCCGTACTCCCATCACC
L R A L R P L R A L S R F E G M R V V V -
TGGATGCCCTCGTGGGCGCCATCCCCTCCATCATGAACGTCCTCCTCGTCTGCCTCATCT
4021
ACCTACGGGAGCACCCGCGGTAGGGGAGGTAGTACTTGCAGGAGGAGCAGACGGAGTAGA
D A L V G A I P S I M N V L L V C L I F -
TCTGGCTCATCTTCAGCATCATGGGCGTGAACCTCTTCGCCGGGAAATTTTCGAAGTGCG
4081
AGACCGAGTAGAAGTCGTAGTACCCGCACTTGGAGAAGCGGCCCTTTAAAAGCTTCACGC
W L I F S I M G V N L F A G K F S K C V -
TCGACACCAGAAATAACCCATTTTCCAACGTGAATTCGACGATGGTGAATAACAAGTCCG
4141
AGCTGTGGTCTTTATTGGGTAAAAGGTTGCACTTAAGCTGCTACCACTTATTGTTCAGGC

```
      D  T  R  N  N  P  F  S  N  V  N  S  T  M  V  N  N  K  S  E -

      AGTGTCACAATCAAAACAGCACCGGCCACTTCTTCTGGGTCAACGTCAAAGTCAACTTCG
4201  ---------+---------+---------+---------+---------+---------+
      TCACAGTGTTAGTTTTGTCGTGGCCGGTGAAGAAGACCCAGTTGCAGTTTCAGTTGAAGC

      C  H  N  Q  N  S  T  G  K  F  F  W  V  N  V  K  V  N  F  D -

      ACAACGTCGCTATGGGCTACCTCGCACTTCTTCAGGTGGCAACCTTCAAAGGCTGGATGG
4261  ---------+---------+---------+---------+---------+---------+
      TGTTGCAGCGATACCCGATGGAGCGTGAAGAAGTCCACCGTTGGAAGTTTCCGACCTACC

      N  V  A  M  G  Y  L  A  L  L  Q  V  A  T  F  K  G  W  M  D -

      ACATAATGTATGCAGCTGTTGATTCCGGAGAGATCAACAGTCAGCCTAACTGGGAGAACA
4321  ---------+---------+---------+---------+---------+---------+
      TGTATTACATACGTCGACAACTAAGGCCTCTCTAGTTGTCAGTCGGATTGACCCTCTTGT

      I  M  Y  A  A  V  D  S  G  E  I  N  S  Q  P  N  W  E  N  N -

      ACTTGTACATGTACCTGTACTTCGTCGTTTTCATCATTTTCGGTGGCTTCTTCACGCTGA
4381  ---------+---------+---------+---------+---------+---------+
      TGAACATGTACATGGACATGAAGCAGCAAAAGTAGTAAAAGCCACCGAAGAAGTGCGACT

      L  Y  M  Y  L  Y  F  V  V  F  I  I  F  G  G  F  F  T  L  N -

      ATCTCTTTGTTGGGGTCATAATCGACAACTTCAACCAACAGAAAAAAAAAGCTAGGAGGCC
4441  ---------+---------+---------+---------+---------+---------+
      TAGAGAAACAACCCCAGTATTAGCTGTTGAAGTTGGTTGTCTTTTTTTTCGATCCTCCGG

      L  F  V  G  V  I  I  D  N  F  N  Q  Q  K  K  K  L  G  G  Q -

      AGGACATCTTCATGACAGAAGAGCAGAAGAAGTACTACAATGCCATGAAGAAGCTGGGCT
4501  ---------+---------+---------+---------+---------+---------+
      TCCTGTAGAAGTACTGTCTTCTCGTCTTCTTCATGATGTTACGGTACTTCTTCGACCCGA

      D  I  F  M  T  E  E  Q  K  K  Y  Y  N  A  M  K  K  L  G  S -

      CCAAGAAACCCCAGAAGCCCATCCCACGGCCCCTGAATAAGTACCAAGGCTTCGTGTTTG
4561  ---------+---------+---------+---------+---------+---------+
      GGTTCTTTGGGGTCTTCGGGTAGGGTGCCGGGGACTTATTCATGGTTCCGAAGCACAAAC

      K  K  P  Q  K  P  I  P  R  P  L  N  K  Y  Q  G  F  V  F  D -

      ACATCGTGACCAGGCAAGCCTTTGACATCATCATCATGGTTCTCATCTGCCTCAACATGA
4621  ---------+---------+---------+---------+---------+---------+
      TGTAGCACTGGTCCGTTCGGAAACTGTAGTAGTAGTACCAAGAGTAGACGGAGTTGTACT

      I  V  T  R  Q  A  F  D  I  I  I  M  V  L  I  C  L  N  M  Y -

      TCACCATGATGGTGGAGACCGACGAGCAGGGCGAGGAGAAGACGAAGGTTCTGGGCAGAA
4681  ---------+---------+---------+---------+---------+---------+
      AGTGGTACTACCACCTCTGGCTGCTCGTCCCGCTCCTCTTCTGCTTCCAAGACCCGTCTT

      T  M  M  V  E  T  D  E  Q  G  E  E  K  T  K  V  L  G  R  Y -

      TCAACCAGTTCTTTGTGGCCGTCTTCACGGGCGAGTGTGTGATGAAGATGTTCGCCCTGC
4741  ---------+---------+---------+---------+---------+---------+
      AGTTGGTCAAGAAACACCGGCAGAAGTGCCCGCTCACACACTACTTCTACAAGCGGGACG

      N  Q  F  F  V  A  V  F  T  G  E  C  V  M  K  M  F  A  L  R -
```

```
     GACAGTACTACTTCACCAACGGCTGGAACGTGTTCGACTTCATAGTGGTGATCCTGTCCA
4801 ---------+---------+---------+---------+---------+---------+
     CTGTCATGATGAAGTGGTTGCCGACCTTGCACAAGCTGAAGTATCACCACTAGGACAGGT

      Q  Y  Y  F  T  N  G  W  N  V  F  D  F  I  V  V  I  L  S  I -

     TTGGGAGTCTGCTGTTTTCTGCAATCCTTAAGTCACTGGAAAACTACTTCTCCCCGACGC
4861 ---------+---------+---------+---------+---------+---------+
     AACCCTCAGACGACAAAAGACGTTAGGAATTCAGTGACCTTTTGATGAAGAGGGGCTGCG

       G  S  L  L  F  S  A  I  L  K  S  L  E  N  Y  F  S  P  T  L -

     TCTTCCGGGTCATCCGTCTGGCCAGGATCGGCCGCATCCTCAGGCTGATCCGAGCAGCCA
4921 ---------+---------+---------+---------+---------+---------+
     AGAAGGCCCAGTAGGCAGACCGGTCCTAGCCGGCGTAGGAGTCCGACTAGGCTCGTCGGT

       F  R  V  I  R  L  A  R  I  G  R  I  L  R  L  I  R  A  A  K -

     AGGGGATTCGCACGCTGCTCTTCGCCCTCATGATGTCCCTGCCCGCCCTCTTCAACATCG
4981 ---------+---------+---------+---------+---------+---------+
     TCCCCTAAGCGTGCGACGAGAAGCGGGAGTACTACAGGGACGGGCGGGAGAAGTTGTAGC

      G  I  R  T  L  L  F  A  L  M  M  S  L  P  A  L  F  N  I  G -

     GCCTCCTCCTCTTCCTCGTCATGTTCATCTACTCCATCTTCGGCATGGCCAGCTTCGCTA
5041 ---------+---------+---------+---------+---------+---------+
     CGGAGGAGGAGAAGGAGCAGTACAAGTAGATGAGGTAGAAGCCGTACCGGTCGAAGCGAT

       L  L  L  F  L  V  M  F  I  Y  S  I  F  G  M  A  S  F  A  N -

     ACGTCGTGGACGAGGCCGGCATCGACGACATGTTCAACTTCAAGACCTTTGGCAACAGCA
5101 ---------+---------+---------+---------+---------+---------+
     TGCAGCACCTGCTCCGGCCGTAGCTGCTGTACAAGTTGAAGTTCTGGAAACCGTTGTCGT

       V  V  D  E  A  G  I  D  D  M  F  N  F  K  T  F  G  N  S  M -

     TGCTGTGCCTGTTCCAGATCACCACCTCGGCCGGCTGGGACGGCCTCCTCAGCCCCATCC
5161 ---------+---------+---------+---------+---------+---------+
     ACGACACGGACAAGGTCTAGTGGTGGAGCCGGCCGACCCTGCCCGAGGAGTCGGGGTAGG

       L  C  L  F  Q  I  T  T  S  A  G  W  D  G  L  L  S  P  I  L -

     TCAACACGGGGCCTCCCTACTGCGACCCCAACCTGCCCAACAGCAACGGCTCCCGGGGGA
5221 ---------+---------+---------+---------+---------+---------+
     AGTTGTGCCCCGGAGGGATGACGCTGGGGTTGGACGGGTTGTCGTTGCCGAGGGCCCCCT

       N  T  G  P  P  Y  C  D  P  N  L  P  N  S  N  G  S  R  G  N -

     ACTGCGGGAGCCCGGCGGTGGGCATCATCTTCTTCACCACCTACATCATCATCTCCTTCC
5281 ---------+---------+---------+---------+---------+---------+
     TGACGCCCTCGGGCCGCCACCCGTAGTAGAAGAAGTGGTGGATGTAGTAGTAGAGGAAGG

       C  G  S  P  A  V  G  I  I  F  F  T  T  Y  I  I  I  S  F  L -

     TCATCGTGGTCAACATGTACATCGCAGTGATTCTGGAGAACTTCAACGTAGCCACCGAGG
5341 ---------+---------+---------+---------+---------+---------+
     AGTAGCACCAGTTGTACATGTAGCGTCACTAAGACCTCTTGAAGTTGCATCGGTGGCTCC

       I  V  V  N  M  Y  I  A  V  I  L  E  N  F  N  V  A  T  E  E -
```

```
      AGAGCACCGAGCCCCCTGAGCGAGGACGACTTCGACATGTTCTATGAGACCTGGGAGAAGT
5401  ---------+---------+---------+---------+---------+---------+
      TCTCGTGCCTCGGGGACTCGCTCCTGCTGAAGCTGTACAAGATACTCTGGACCCTCTTCA

       S  T  E  P  L  S  E  D  D  F  D  M  F  Y  E  T  W  E  K  F -

      TCGACCCGGAGGCCACCCAGTTCATTGCCTTTTCTGCCCTCTCAGACTTCGCGGACACGC
5461  ---------+---------+---------+---------+---------+---------+
      AGCTGGGCCTCCGGTGGGTCAAGTAACGGAAAAGACGGGAGAGTCTGAAGCGCCTGTGCG

       D  P  E  A  T  Q  F  I  A  F  S  A  L  S  D  F  A  D  T  L -

      TCTCCGGCCCTCTTAGAATCCCCAAACCCAACCAGAATATATTAATCCAGATGGACCTGC
5521  ---------+---------+---------+---------+---------+---------+
      AGAGGCCGGGAGAATCTTAGGGGTTTGGGTTGGTCTTATATAATTAGGTCTACCTGGACG

       S  G  P  L  R  I  P  K  P  N  Q  N  I  L  I  Q  M  D  L  P -

      CGTTGGTCCCCGGGGATAAGATCCACTGTCTGGACATCCTTTTTGCCTTCACAAAGAACG
5581  ---------+---------+---------+---------+---------+---------+
      GCAACCAGGGGCCCCTATTCTAGGTGACAGACCTGTAGGAAAAACGGAAGTGTTTCTTGC

       L  V  P  G  D  K  I  H  C  L  D  I  L  F  A  F  T  K  N  V -

      TCTTGGGAGAATCCGGGGAGTTGGACTCCCTGAAGACCAATATGGAAGAGAAGTTTATGG
5641  ---------+---------+---------+---------+---------+---------+
      AGAACCCTCTTAGGCCCCTCAACCTGAGGGACTTCTGGTTATACCTTCTCTTCAAATACC

       L  G  E  S  G  E  L  D  S  L  K  T  N  M  E  E  K  F  M  A -

      CGACCAATCTCTCCAAAGCATCCTATGAACCAATAGCCACCACCCTCCGGTGGAAGCAGG
5701  ---------+---------+---------+---------+---------+---------+
      GCTGGTTAGAGAGGTTTCGTAGGATACTTGGTTATCGGTGGTGGGAGGCCACCTTCGTCC

       T  N  L  S  K  A  S  Y  E  P  I  A  T  T  L  R  W  K  Q  E -

      AAGACCTCTCAGCCACAGTCATTCAAAAGGCCTACCGGAGCTACATGCTGCACCGCTCCT
5761  ---------+---------+---------+---------+---------+---------+
      TTCTGGAGAGTCGGTGTCAGTAAGTTTTCCGGATGGCCTCGATGTACGACGTGGCGAGGA

       D  L  S  A  T  V  I  Q  K  A  Y  R  S  Y  M  L  H  R  S  L -

      TGACACTCTCCAACACCCTGCATGTGCCCAGGGCTGAGGAGGATGGCGTGTCACTTCCCG
5821  ---------+---------+---------+---------+---------+---------+
      ACTGTGAGAGGTTGTGGGACGTACACGGGTCCCGACTCCTCCTACCGCACAGTGAAGGGC

       T  L  S  N  T  L  H  V  P  R  A  E  E  D  G  V  S  L  P  G -

      GGGAAGGCTACATTACATTCATGGCAAACAGTGGACTCCCGGACAAATCAGAAACTGCCT
5881  ---------+---------+---------+---------+---------+---------+
      CCCTTCCGATGTAATGTAAGTACCGTTTGTCACCTGAGGGCCTGTTTAGTCTTTGACGGA

       E  G  Y  Y  T  F  M  A  N  S  G  L  P  D  K  S  E  T  A  S -

      CTGCTACGTCTTTCCCGCCATCCTATGACAGTGTCACCAGGGGCCTGAGTGACCGGGCCA
5941  ---------+---------+---------+---------+---------+---------+
      GACGATGCAGAAAGGGCGGTAGGATACTGTCACAGTGGTCCCCGGACTCACTGGCCCGGT

       A  T  S  F  P  P  S  Y  D  S  V  T  R  G  L  S  D  R  A  N -

      ACATTAACCCATCTAGCTCAATGCAAAATGAAGATGAGGTCGCTGCTAAGGAAGGAAACA
6001  ---------+---------+---------+---------+---------+---------+
```

```
TGTAATTGGGTACATCGAGTTACGTTTTACTTCTACTCCAGCGACGATTCCTTCCTTTGT

 I  N  P  S  S  S  M  Q  N  E  D  E  V  A  A  K  E  G  N  S -

GCCCTGGACCTCAGTGAAGGCACTCAGGCATGCACAGGGCAGGTTCCAATGTCTTTCTCT
---------+---------+---------+---------+---------+---------+
CGGGACCTGGAGTCACTTCCGTGAGTCCGTACGTGTCCCGTCCAAGGTTACAGAAAGAGA

  P  G  P  Q  *  R  H  S  G  M  H  R  A  G  S  N  V  F  L  C -

GCTGTACTAACTCCTTCCCTCTGGAGGTGGCACCAACCTCCAGCCTCCACCAATGCATGT
---------+---------+---------+---------+---------+---------+
CGACATGATTGAGGAAGGGAGACCTCCACCGTGGTTGGAGGTCCGAGCTGGTTACGTACA

 C  T  N  S  F  P  L  E  V  A  P  T  S  S  L  H  Q  C  M  S -

CACTGGTCATGGTGTCAGAACTGAATGGGGACATCCTTGAGAAAGCCCCCACCCCAATAG
---------+---------+---------+---------+---------+---------+
GTGACCAGTACCACAGTCTTGACTTACCCCTGTAGGAACTCTTTCGGGGGTGGGGTTATC

 L  V  M  V  S  E  L  N  G  D  I  L  E  K  A  P  T  P  I  G -

GAATCAAAAGCCAAGGATACTCCTCCATTCTGACGTCCCTTCCGAGTTCCCAGAAGATGT
---------+---------+---------+---------+---------+---------+
CTTAGTTTTCGGTTCCTATGAGGAGGTAAGACTGCAGGGAAGGCTCAAGGGTCTTCTACA

 I  K  S  Q  G  Y  S  S  I  L  T  S  L  P  S  S  Q  K  M  S -

CATTGCTCCCTTCTGTTTGTGACCAGAGACGTGATTCACCAACTTCTCGGAGCCAGAGAC
---------+---------+---------+---------+---------+---------+
GTAACGAGGGAAGACAAACACTGGTCTCTGCACTAAGTGGTTGAAGAGCCTCGGTCTCTG

  L  L  P  S  V  C  D  Q  R  R  D  S  P  T  S  R  S  Q  R  H -

ACATAGCAAAGACTTTTCTGCTGGTGTCGGGCAGTCTTAGAGAAGTCACGTAGGGGTTGG
---------+---------+---------+---------+---------+---------+
TGTATCGTTTCTGAAAAGACGACCACAGCCCGTCAGAATCTCTTCAGTGCATCCCCAACC

 I  A  K  T  F  L  L  V  S  G  S  L  R  E  V  T  *

TACTGAGAATTAGGGTTTGCATGACTGCATGCTCACAGCTGCCGGACAATACCTGTGAGT
---------+---------+---------+---------+---------+---------+
ATGACTCTTAATCCCAAACGTACTGACGTACGAGTGTCGACGGCCTGTTATGGACACTCA


CGGCCATTAAAATTAATATTTTTAAAGTTAAAAAAAAAAAAAAAA
---------+---------+---------+---------+---- 6524
GCCGGTAATTTTAATTATAAAAATTTCAATTTTTTTTTTTTTTTT
```

Figure 1c

Nucleic acid and amino acid sequence of variant TTXi DRG sodium channel - SNS-short form (same numbering as SNS-B)

5' ctgggagagaaagcgtctcgcctagcgactcccagagctttaagccgggaagggacaagc

gtcaggacatctcagaatcccgaaccttctagggagggaggttcttacctccatgcttcc

cgtaggaacctaatcccaattatttagctgtatttataatacaaaatatgaatgtta

aatgtacaaaatgctttcccagcatgcctgcatctcctcctagagtcctgttcccaagcc

ctctctactctcagtactgtagaaaagaaataagctttacgtgagaaacccaggcactgg

atcttatccaggtgctcacctcagagtctttagtgggtgtagcgctgtggtagagcattt

ggttatagatacaaacccagggcagggagactgcagtggccattctctcccaggccagac

gtgccctgatccttcccacagagatgagaaggctggaaccagaacactcaggttttggct

tctcttgggggaggagaggtaatcttgttactttaataacatcagtgtgtcccctcctctac

taggaggccaggacatcttc

```
                ATGACAGAAGAGCAGAAGAAGTACTACAATGCCATGAAGAAGCTGGGCT
                --------+---------+---------+---------+---------+
                TACTGTCTTCTCGTCTTCTTCATGATGTTACGGTACTTCTTCGACCCGA

                M  T  E  E  Q  K  K  Y  Y  N  A  M  K  K  L  G  S  -

                CCAAGAAACCCCAGAAGCCCATCCCACGGCCCCTGAATAAGTACCAAGCGTTCGTGTTTG
4561            ---------+---------+---------+---------+---------+---------+
                GGTTCTTTGGGGTCTTCGGGTAGGGTGCCGGGGACTTATTCATGGTTCCGAAGCACAAAC

                K  K  P  Q  K  P  I  P  R  P  L  N  K  Y  Q  G  F  V  F  D  -

                ACATCGTGACCAGGCAAGCCTTTGACATCATCATCATGGTTCTCATCTGCCTCAACATGA
4621            ---------+---------+---------+---------+---------+---------+
                TGTAGCACTGGTCCGTTCGGAAACTGTAGTAGTAGTACCAAGAGTAGACGGAGTTGTACT

                 I  V  T  R  Q  A  F  D  I  I  I  M  V  L  I  C  L  N  M  I  -

                TCACCATGATGGTGGAGACCGACGAGCAGGCCGAGGAGAAGACGAAGGTTCTGGGCAGAA
4681            ---------+---------+---------+---------+---------+---------+
                AGTGGTACTACCACCTCTGGCTGCTCGTCCCGCTCCTCTTCTGCTTCCAAGACCCGTCTT

                 T  M  M  V  E  T  D  E  Q  G  E  E  K  T  K  V  L  G  R  I  -

                TCAACCAGTTCTTTGTGGCCGTCTTCACGGGCGAGTGTGTGATGAAGATGTTCGCCCTGC
4741            ---------+---------+---------+---------+---------+---------+
                AGTTGGTCAAGAAACACCGGCAGAAGTGCCCGCTCACACACTACTTCTACAAGCGGGACG

                 N  Q  F  F  V  A  V  F  T  G  E  C  V  M  K  M  F  A  L  R  -

                GACAGTACTACTTCACCAACGGCTGGAACGTGTTCGACTTCATAGTGGTGATCCTGTCCA
4801            ---------+---------+---------+---------+---------+---------+
                CTGTCATGATGAAGTGGTTGCCGACCTTGCACAAGCTGAAGTATCACCACTAGGACAGGT
```

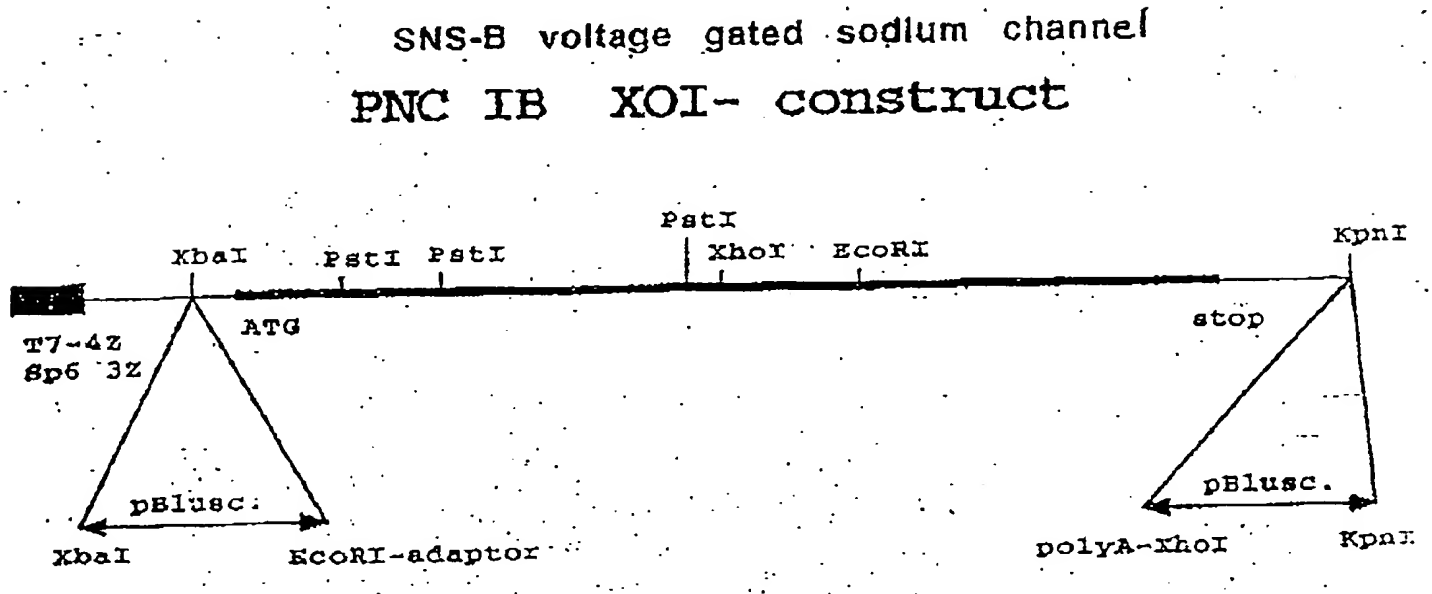
```
      Q  Y  Y  F  T  N  G  W  N  V  F  D  F  I  V  V  I  L  S  I -

      TTGGGAGTCTGCTGTTTTCTGCAATCCTTAAGTCACTGGAAAACTACTTCTCCCCGACGC
4861  ---------+---------+---------+---------+---------+---------+
      AACCCTCAGACGACAAAAGACGTTAGGAATTCAGTGACCTTTTGATGAAGAGGGGCTGCG

        G  S  L  L  F  S  A  I  L  K  S  L  E  N  Y  F  S  P  T  L -

      TCTTCCGGGTCATCCGTCTGGCCAGGATCGGCCGCATCCTCAGGCTGATCCGAGCAGCCA
4921  ---------+---------+---------+---------+---------+---------+
      AGAAGGCCCAGTAGGCAGACCGGTCCTAGCCGGCGTAGGAGTCCGACTAGGCTCGTCGGT

        F  R  V  I  R  L  A  R  I  G  R  I  L  R  L  I  R  A  A  K -

      AGGGGATTCGCACGCTGCTCTTCGCCCTCATGATGTCCCTGCCCGCCCTCTTCAACATCG
4981  ---------+---------+---------+---------+---------+---------+
      TCCCCTAAGCGTGCGACGAGAAGCGCGAGTACTACAGGGACGGGCGCGAGAAGTTGTAGC

        G  I  R  T  L  L  F  A  L  M  M  S  L  P  A  L  F  N  I  G -

      GCCTCCTCCTCTTCCTCGTCATGTTCATCTACTCCATCTTCGGCATGGCCAGCTTCGCTA
5041  ---------+---------+---------+---------+---------+---------+
      CGGAGGAGGAGAAGGAGCAGTACAAGTAGATGAGGTAGAAGCCGTACCGGTCGAAGCGAT

        L  L  L  F  L  V  M  F  I  Y  S  I  F  G  M  A  S  F  A  N -

      ACGTCGTGGACGAGGCCGGCATCGACGACATGTTCAACTTCAAGACCTTTGGCAACAGCA
5101  ---------+---------+---------+---------+---------+---------+
      TGCAGCACCTGCTCCGGCCGTAGCTGCTGTACAAGTTGAAGTTCTGGAAACCGTTGTCGT

        V  V  D  E  A  G  I  D  D  M  F  N  F  K  T  F  G  N  S  M -


      TGCTGTGCCTGTTCCAGATCACCACCTCGGCCGGCTGGGACGGCCTCCTCAGCCCCATCC
5161  ---------+---------+---------+---------+---------+---------+
      ACGACACGGACAAGGTCTAGTGGTGGAGCCGGCCGACCCTGCCGGAGGAGTCGGGGTAGG

        L  C  L  F  Q  I  T  T  S  A  G  W  D  G  L  L  S  P  I  L -

      TCAACACGGGGCCTCCCTACTGCGACCCCAACCTGCCCAACAGCAACGGCTCCCGGGGGA
5221  ---------+---------+---------+---------+---------+---------+
      AGTTGTGCCCCGGAGGGATGACGCTGGGGTTGGACGGGTTGTCGTTGCCGAGGGCCCCCT

        N  T  G  P  P  Y  C  D  P  N  L  P  N  S  N  G  S  R  G  N -

      ACTGCGGGAGCCCGGCGGTGGGCATCATCTTCTTCACCACCTACATCATCATCTCCTTCC
5281  ---------+---------+---------+---------+---------+---------+
      TGACGCCCTCGGGCCGCCACCCGTAGTAGAAGAAGTCGTGGATGTAGTAGTAGACGAAGG

        C  G  S  P  A  V  G  I  I  F  F  T  T  Y  I  I  I  S  F  L -

      TCATCGTGGTCAACATGTACATCGCAGTGATTCTGGAGAACTTCAACGTAGCCACCGAGG
5341  ---------+---------+---------+---------+---------+---------+
      AGTAGCACCAGTTGTACATGTAGCGTCACTAAGACCTCTTGAAGTTGCATCGGTGGCTCC

        I  V  V  N  M  Y  I  A  V  I  L  E  N  F  N  V  A  T  E  E -

      ACAGCACGGAGCCCCTGAGCGAGGACGACTTCGACATGTTCTATGAGACCTGGGAGAAGT
5401  ---------+---------+---------+---------+---------+---------+
      TCTCGTGCCTCGGGGACTCGCTCCTGCTGAAGCTGTACAAGATACTCTGGACCCTCTTCA
```

```
      S  T  E  P  L  S  E  D  D  F  D  M  F  Y  E  T  W  E  K  F -

      TCGACCCGGAGGCCACCCAGTTCATTGCCTTTTCTGCCCTCTCAGACTTCGCGGACACGC
5461  ---------+---------+---------+---------+---------+---------+
      AGCTGGGCCTCCGGTGGGTCAAGTAACGGAAAAGACGGGAGAGTCTGAAGCGCCTGTGCG

        D  P  E  A  T  Q  F  I  A  F  S  A  L  S  D  F  A  D  T  L -

      TCTCCGGCCCTCTTAGAATCCCCAAACCCAACCAGAATATATTAATCCAGATGGACCTGC
5521  ---------+---------+---------+---------+---------+---------+
      AGAGGCCGGGAGAATCTTAGGGGTTTGGGTTGGTCTTATATAATTAGGTCTACCTGGACG

    S  G  P  L  R  I  P  K  P  N  Q  N  I  L  I  Q  M  D  L  P -

      CGTTGGTCCCCGGGGATAAGATCCACTGTCTGGACATCCTTTTTGCCTTCACAAAGAACG
5581  ---------+---------+---------+---------+---------+---------+
      GCAACCAGGGGCCCCTATTCTAGGTGACAGACCTGTAGGAAAAACGGAAGTGTTTCTTGC

        L  V  P  G  D  K  I  H  C  L  D  I  L  F  A  F  T  K  N  V -

      TCTTGGGAGAATCCGGGGAGTTGGACTCCCTGAAGACCAATATGGAAGAGAAGTTTATGG
5641  ---------+---------+---------+---------+---------+---------+
      AGAACCCTCTTAGGCCCCTCAACCTGAGGGACTTCTGGTTATACCTTCTCTTCAAATACC

        L  G  E  S  G  E  L  D  S  L  K  T  N  M  E  E  K  F  M  A -

      CGACCAATCTCTCCAAAGCATCCTATGAACCAATAGCCACCACCCTCCGGTGGAAGCAGG
5701  ---------+---------+---------+---------+---------+---------+
      GCTGGTTAGAGAGGTTTCGTAGGATACTTGGTTATCGGTGGTGGGAGGCCACCTTCGTCC

        T  N  L  S  K  A  S  Y  E  P  I  A  T  T  L  R  W  K  Q  E -

      AAGACCTCTCAGCCACAGTCATTCAAAAGGCCTACCGGAGCTACATGCTGCACCGCTCCT
5761  ---------+---------+---------+---------+---------+---------+
      TTCTGGAGAGTCGGTGTCAGTAAGTTTTCCGGATGGCCTCGATGTACGACGTGGCGAGGA

        D  L  S  A  T  V  I  Q  K  A  Y  R  S  Y  M  L  H  R  S  L -

      TGACACTCTCCAACACCCTGCATGTGCCCAGGGCTGAGGAGGATGGCGTGTCACTTCCCG
5821  ---------+---------+---------+---------+---------+---------+
      ACTGTGAGAGGTTGTGGGACGTACACGGGTCCCGACTCCTCCTACCGCACAGTGAAGGGC

        T  L  S  N  T  L  H  V  P  R  A  E  E  D  G  V  S  L  P  G -

      GGGAAGGCTACATTACATTCATGGCAAACAGTGGACTCCCGGACAAATCAGAAACTGCCT
5881  ---------+---------+---------+---------+---------+---------+
      CCCTTCCGATGTAATGTAAGTACCGTTTGTCACCTGAGGGCCTGTTTAGTCTTTGACGGA

        E  G  Y  I  T  F  M  A  N  S  G  L  P  D  K  S  E  T  A  S -

      CTGCTACGTCTTTCCCGCCATCCTATGACAGTGTCACCAGGGGCCTGAGTGACCGGGCCA
5941  ---------+---------+---------+---------+---------+---------+
      GACGATGCAGAAAGGGCGGTAGGATACTGTCACAGTCGTCCCCGGACTCACTGGCCCCGT

        A  T  S  F  P  P  S  Y  D  S  V  T  R  G  L  S  D  R  A  N -

      ACATTAACCCATCTAGCTCAATGCAAAATGAAGATGAGGTCGCTGCTAAGGAAGGAAACA
6001  ---------+---------+---------+---------+---------+---------+
      TGTAATTGGGTAGATCGAGTTACGTTTTACTTCTACTCCAGCGACGATTCCTTCCTTTGT

      I  N  P  S  S  S  M  Q  N  E  D  E  V  A  A  K  E  G  N  S -
```

```
     GCCCTGGACCTCACTGAAGGCACTCAGGCATGCACAGGGCAGGTTCCAATGTCTTTCTCT
6061 ---------+---------+---------+---------+---------+---------+
     CGGGACCTGGAGTCACTTCCGTGAGTCCGTACGTGTCCCGTCCAAGGTTACAGAAAGAGA

       P  G  P  Q  *  R  H  S  G  M  H  R  A  G  S  N  V  F  L  C -

     GCTGTACTAACTCCTTCCCTCTGGAGGTGGCACCAACCTCCAGCCTCCACCAATGCATGT
6121 ---------+---------+---------+---------+---------+---------+
     CGACATGATTGAGGAAGGGAGACCTCCACCGTGGTTGGAGGTCGGAGGTGGTTACGTACA

       C  T  N  S  F  P  L  E  V  A  P  T  S  S  L  H  Q  C  M  S -

     CACTGGTCATGGTGTCAGAACTGAATGGGGACATCCTTGAGAAAGCCCCCACCCCAATAG
6181 ---------+---------+---------+---------+---------+---------+
     GTGACCAGTACCACAGTCTTGACTTACCCCTGTAGGAACTCTTTCGGGGGTGGGGTTATC

       L  V  M  V  S  E  L  N  G  D  I  L  E  K  A  P  T  P  I  G -

     GAATCAAAAGCCAAGGATACTCCTCCATTCTGACGTCCCTTCCGAGTTCCCAGAAGATGT
6241 ---------+---------+---------+---------+---------+---------+
     CTTAGTTTTCGGTTCCTATGAGGAGGTAAGACTGCAGGGAAGGCTCAAGGGTCTTCTACA

       I  K  S  Q  G  Y  S  S  I  L  T  S  L  P  S  S  Q  K  M  S -

     CATTGCTCCCTTCTGTTTGTGACCAGAGACGTGATTCACCAACTTCTCGGAGCCAGAGAC
6301 ---------+---------+---------+---------+---------+---------+
     GTAACGAGGGAAGACAAACACTGGTCTCTGCACTAAGTGGTTGAAGAGCCTCGGTCTCTG

       L  L  P  S  V  C  D  Q  R  R  D  S  P  T  S  R  S  Q  R  H -

     ACATAGCAAAGACTTTTCTGCTGGTGTCGGGCAGTCTTAGAGAAGTCACGTAGGGGTTGG
6361 ---------+---------+---------+---------+---------+---------+
     TGTATCGTTTCTGAAAAGACGACCACAGCCCGTCAGAATCTCTTCAGTGCATCCCCAACC

       I  A  K  T  F  L  L  V  S  G  S  L  R  E  V  T  *        --

     TACTGAGAATTAGGGTTTGCATGACTGCATGCTCACAGCTGCCGGACAATACCTGTGAGT
6421 ---------+---------+---------+---------+---------+---------+
     ATGACTCTTAATCCCAAACGTACTGACGTACGAGTGTCGACGGCCTGTTATGGACACTCA


     CGGCCATTAAAATTAATATTTTTAAAGTTAAAAAAAAAAAAAAA
6481 ---------+---------+---------+---------+---- 6524
     GCCGGTAATTTTAATTATAAAAATTTCAATTTTTTTTTTTTTT
```

Figure 1d

Structure of SNS-B voltage-gated sodium channel in pGEM-3Z



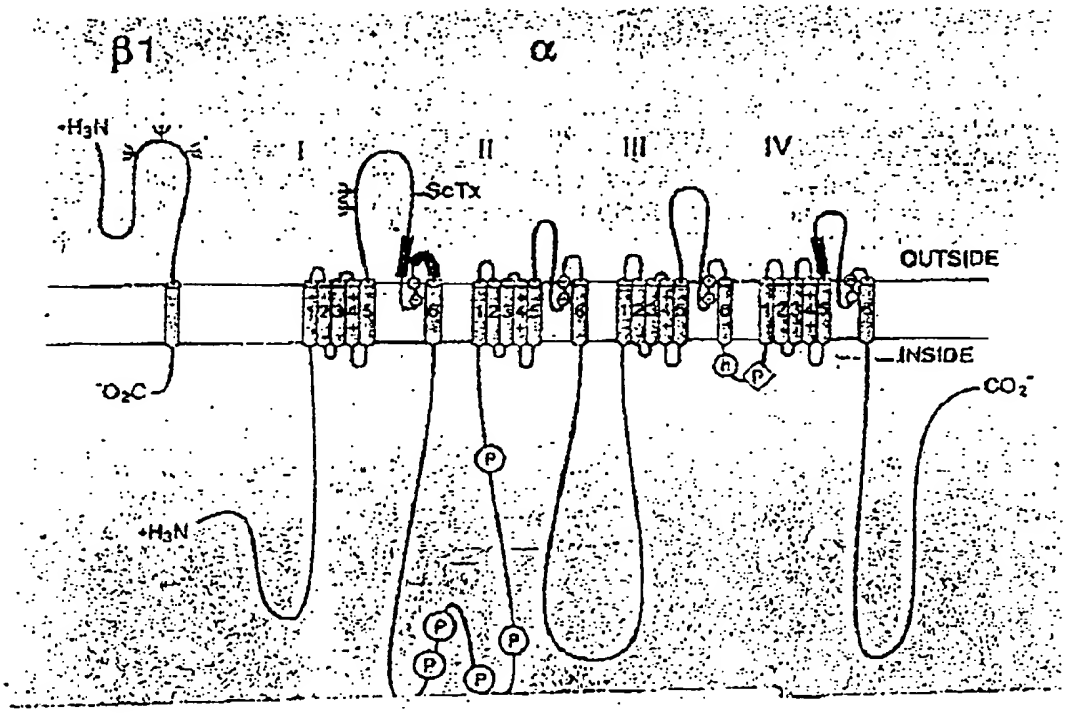
Constructs were generated in pGem 3Z

nd pGem 4Z with bluescript polylinkers

. Linearization site is KPNI

Figure 1e

Schematised drawing of voltage-gated sodium channel (from Caterall 1992)



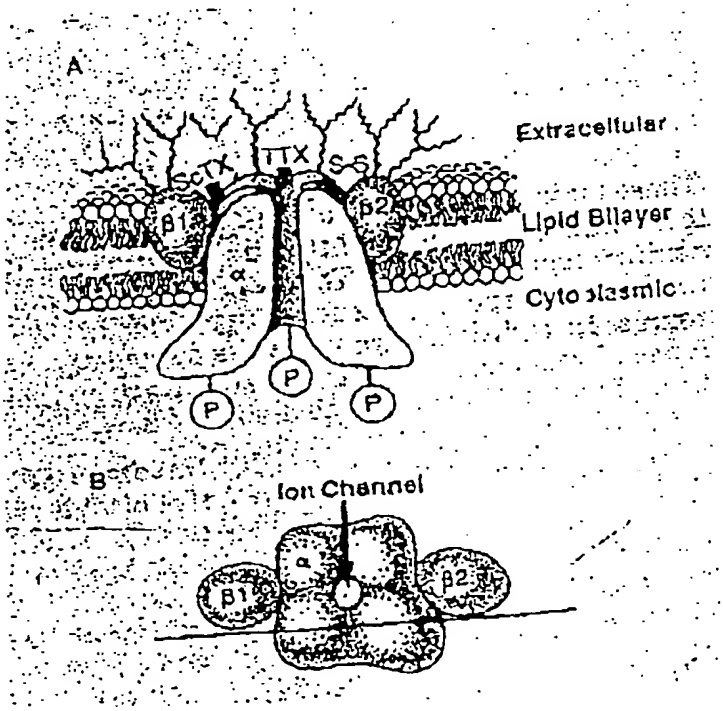


Figure 2

Sequence of PCR primers for isolation of human clone probes

a) Highly conserved regions of all sodium channels

1) Position 2475- 2510 S4 Domain 11

Degenerate primers (20-24mers) encoding amino acid residues

RLLRVFKLAKSWPTL

or non degenerate primers within this region

e.g. 5' gcttgctgcgggtcttcaagc 3'

2) Position 3961 - 4010 S4 domain III

Degenerate primers encoding the complementary strand encoding residues

LRALPLRALSRFEG or non degenerate primers within this region

e.g. 5' atcgagacagagcccgcagcg 3'

b) Unique sequence primers for SNS-homologues

e.g. residues within the region 2641-2680

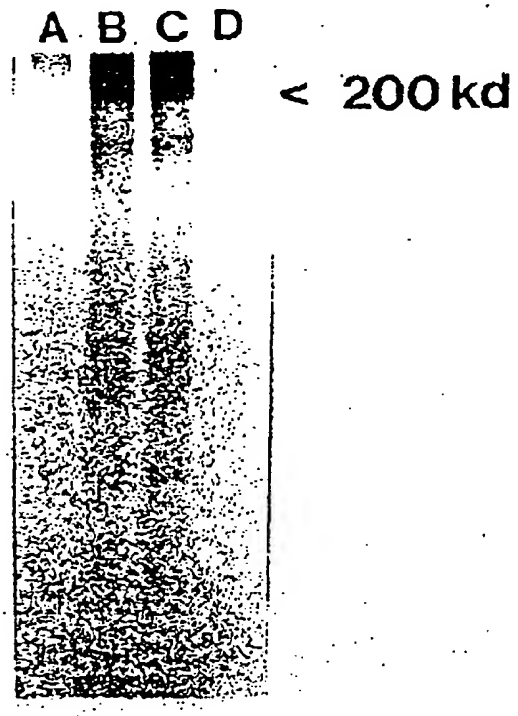
e.g. 5' acgggtgccgcaaggacggcgtctccgtgtggaacggcgagaag 3'

and complementary sequence within the region 3375 and 3420

e.g. 5' ggctatccttcctcttccagctctcacccaggtatggagccaggt 3'

Figure 3

In vitro synthesis of S-35 methionine labelled SNS-B voltage gated sodium channel in a coupled transcription/translation system



Autoradiograph of a 7.5% SDS polyacrylamide gel, showing the migration of labelled proteins compared to the sizes of known molecular weight markers (Amersham rainbow markers). Lane A control, Lane B SNS-B, Lane C SNS-B, Lane D control. The predicted 200kDa band representing the SNS-B sodium channel is arrowed.

Figure 4a

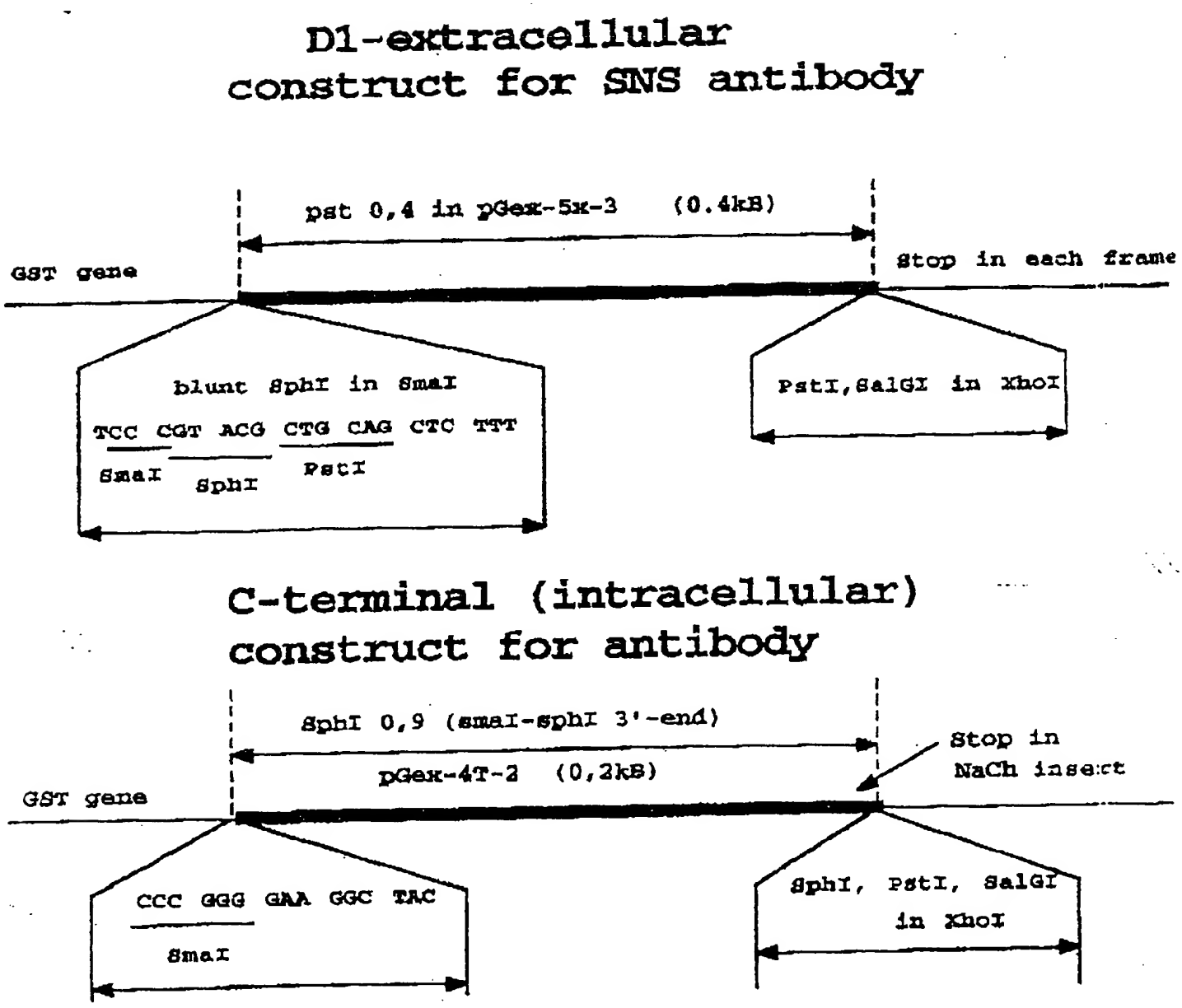
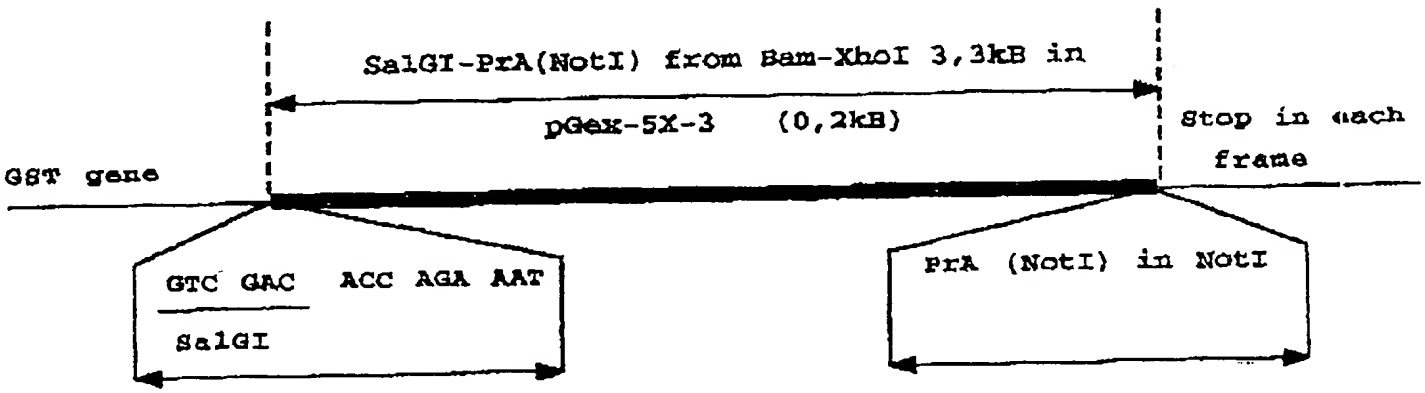

D1-extracellular
construct for SNS antibody
pst 0,4 in pGex-5x-3 (0.4kB)
GST gene
Stop in each frame
blunt SphI in SmaI
TCC CGT ACG CTG CAG CTC TTT
SmaI
SphI
PstI
PstI,SalGI in XhoI
C-terminal (intracellular)
construct for antibody
SphI 0,9 (smaI-sphI 3'-end)
pGex-4T-2 (0,2kB)
Stop in
NaCh insert
GST gene
CCC GGG GAA GGC TAC
SmaI
SphI, PstI, SalGI
in XhoI

Figure 4b

Extracellular D3 construct for antibody



Intracellular D1-D2 construct for antibody